University of Alberta

Population Structure, Diversity and Gene Flow in *Pinus contorta* Dougl. From RAPD Markers and Sequencing Data

by

©

Aron James Fazekas

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in

Conservation Biology

Department of Renewable Resources

Edmonton, Alberta

Fall 2003

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisisitons et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-87969-0
Our file Notre référence
ISBN: 0-612-87969-0

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

Canadä

University of Alberta

Library Release Form

Name of Author: Aron James Fazekas

Title of Thesis: Population Structure, Diversity and Gene Flow in *Pinus contorta* Dougl. From RAPD Markers and Sequencing Data

Degree: Doctor of Philosophy

Year this Degree Granted: 2003

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Date submitted to the Faculty of Graduate Studies and Research: Sept 23/2003

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Population Structure, Diversity and Gene Flow in *Pinus contorta* Dougl. From RAPD Markers and Sequencing Data, submitted by Aron James Fazekas in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Conservation Biology.

Dr. Francis C. Yeh (Supervisor)

Dr. S. Ellen MacDonald

Dr. Sean W. Graham

Dr. Ross Hodgetts

Dr. Jean Bousquet (External Examiner)

Date: Sept 18/2003

For my parents who always encouraged learning,
and Patricia,
who has provided support and encouragement in everything.

Abstract

Lodgepole pine (*Pinus contorta* Dougl.) is an important component of forest ecosystems in western North America. Its current range extends from the Yukon through British Columbia and western Alberta, south to California and southeast along the Rocky Mountains into Idaho, Utah and Colorado. The population structure, genetic diversity, gene flow and relationships among the subspecies of lodgepole pine were investigated using RAPD markers and sequencing data.

Fifteen populations of lodgepole pine were surveyed for diversity across 52 Random Amplified Polymorphic DNAs (RAPDs). The objective was to compare single-locus and multilocus structures in four marginal, three intermediate and eight central populations. Single-locus estimates indicated average observed and expected heterozygosity to be 0.19 and 0.17, respectively. When these estimates were split into population categories, a clear trend of increasing diversity was detected in the direction of marginal to central populations. *F*-statistics indicated an excess of heterozygotes, with F_{IS} ranging from –0.08 for marginal populations to –0.13 for central populations. The estimates of F_{ST} decreased towards the margins of the species range, indicating increased population differentiation. Multilocus analysis showed significant two-locus and high order gametic disequilibria in all 15 populations. The most prominent components of the two-locus analysis were the variance of disequilibrium (46.2%) and the multilocus Wahlund's effect (31.9%).

To explore the relationship among the subspecies of lodgepole pine and the possibility of northern glacial refugia for subsp. *latifolia* and coastal refugia for subsp. *contorta*, 31 populations were analysed using RAPD markers, to determine whether a pattern of isolation by distance exists. Heterozygosity estimates ranged from 0.15 for subsp. *latifolia* to 0.08 for subspecies *bolanderi*. Estimates of G_{ST} for subspecies *latifolia*, *murrayana* and *contorta* were

0.067, 0.036 and 0.079 respectively. A significant pattern of isolation by distance was detected when all 31 populations were analysed together, and for the 19 populations of subsp. *latifolia*.

Sequence data from two chloroplast and one mitochondrial loci were used to evaluate the relationships between the four subspecies of lodgepole pine. At these loci however, nucleotide variation was confined to a small number of individuals and differentiation between subspecies was not detectable. Estimates of mean diversity were 0.000,178 and 0.000,186 for the two chloroplast loci.

Acknowledgements

I would like to thank my supervisory committee, Francis Yeh, Ellen Macdonald, and Sean Graham for their help and assistance with my research project. I am indebted to Changxi Li, Mary Aleksiuk, Andrea Marciowitz, Pamela Gullekson and Renate Meuser for their assistance in the lab. I thank Gabriel Wolken for his help with geographic information software, and Rong-Cai Yang for his assistance with statistical analysis and for reviewing portions of the thesis. The seeds and needle tissue used in this project were provided by or collected with the assistance of the British Columbia Ministry of Forests. I also thank Al Gordon who piqued my curiosity enough to start me down this path.

Funding for this research was provided by the Natural Sciences and Engineering Research Council (NSERC, A2282) to Francis Yeh and by a University of Alberta, Department of Biological Sciences, Challenge Grant in Biodiversity to Aron Fazekas.

Table of Contents

Chapter 1 Introduction 1
    Introduction 1
    Lodgepole pine 2
        Distribution and ecology 2
        Systematics 4
        Historical reconstruction 5
        Subspecies *bolanderi* 6
        Subspecies *contorta* 7
        Subspecies *murrayana* 9
        Subspecies *latifolia* 11
    Evolutionary forces 14
    RAPD markers 20
    Sequence data 21
    Objectives 22
    References 24
Chapter 2 Random amplified polymorphic DNA diversity of marginal and central populations in *Pinus contorta* subsp. *latifolia* 35
    Introduction 35
    Materials and methods 38
        Population sampling 38
        DNA extraction and amplification 39
        Data Analysis 41
    Results 43
        RAPD Variation Within Populations 43
        RAPD Differentiation Among Populations 44
    Discussion 45
    References 53
Chapter 3 Inferences about the post glacial history of *Pinus contorta* from gene flow 69
    Introduction 69
    Materials and methods 72
        Population sampling 72

DNA extraction and amplification 73
Data Analysis 74
Results 77
Discussion 80
Subspecies *latifolia* 85
Subspecies *contorta* 89
Subspecies *murrayana* 92
Conclusion 93
References 95
Chapter 4 Organellar sequence diversity in *Pinus contorta* Dougl. 115
Introduction 115
Materials and methods 118
Population sampling 118
DNA extraction 119
Amplification and sequencing 120
Data analysis 121
Results 122
Discussion 123
Conclusion 130
References 131
Chapter 5 General Conclusions 145
Introduction 145
Research overview 145
Directions for future research 148
References 150

List of Tables

Table 2-1: Sequences of the University of British Columbia (UBC) Biotechnology Laboratory decamer primers used in this study of *Pinus contorta* subsp. *latifolia*. 61

Table 2-2: Frequencies of 33 Polymorphic RAPDs in 15 populations of *Pinus contorta* subsp. *latifolia*. 62

Table 2-3: Comparison of genetic variation and fixation indices in marginal, intermediate and central populations of *Pinus contorta* subsp. *latifolia*. 63

Table 2-4: Comparison of *F*-statistics and number of migrants (*Nm*) across marginal, intermediate, and central populations in *Pinus contorta* subsp. *latifolia*........ 64

Table 2-5: The observed variance (M2), third moment (M3) and fourth moment (M4) of number of heterozygous loci between two randomly chosen gametes *Pinus contorta* subsp. *latifolia*. Values exceeding 95% confidence limits are denoted by *. 65

Table 2-6: Components of variance of multilocus associations in *Pinus contorta* subsp. *latifolia*. 66

Table 3-1: Seed organization and number of seeds analyzed for the 31 populations of *Pinus contorta* used in this study. 101

Table 3-2: Observed and estimated heterozygosity values for 31 populations of lodgepole pine. 102

Table 3-3: Mean genetic distance between and within (diagonal) subspecies of *Pinus contorta*. 103

Table 3-4: Non neutral[†] loci in *Pinus contorta* and its subspecies. 104

Table 3-5: Means and ranges of $\hat{M}$ (genetic similarity) and *d* (geographic distance) for pairwise comparisons all of populations and for populations within each of the three subspecies of *Pinus contorta* (see Figure 3-6 for a graphical representation of pairwise comparisons). 105

Table 3-6: Estimates of slope (*b*) from the regression of log($\hat{M}$) (genetic similarity) on log (*d*) (geographic distance) with lower (L_b) and upper (U_b) limits generated by bootstrapping. 106

Table 4-1: Number of trees sampled per population for 31 populations of *Pinus contorta*. 136

Table 4-2: Sequences of the amplification and sequencing primers used in this study. 137

Table 4-3: Relative positions of single nucleotide polymorphisms and a 5 bp indel in the *trnL* intron in 9 populations of lodgepole pine. The numbers above indicate the position relative to the start of the inton. 138

Table 4-4: Relative positions of single nucleotide polymorphisms, a 1 bp indel, and a 26 bp indel in the *trnL/F* spacer in 7 populations of lodgepole pine. The numbers above indicate the position relative to the start of the spacer. 139

Table 4-5: Evolutionary distance within (diagonal) and between the four subspecies of lodgepole pine based on the chloroplast *trnL* intron and the *trnF* spacer using the method of Tajima and Nei (1984). 140

List of Figures

Figure 1-1: Approximate distribution of the four subspecies of *Pinus contorta*. 34

Figure 2-1: Locations of 15 Populations of *Pinus contorta* subsp. *latifolia*. 67

Figure 2-2: RAPD profiles amlified with UBC primer 635. Lane 1 is the DNA molecular weight marker VI (Boehringer Mannheim) with size in base pairs indicated on the left. Lanes 2–31 are amplified fragments of megagametophyte template DNAs. The arrows on the right represent scored fragments, with their molecular sizes in base pairs. 68

Figure 3-1: Location of 31 populations of *Pinus contorta* used in this study. 107

Figure 3-2: Unrooted neighbour-joining tree of all 31 populations of *Pinus contorta* used in this study, based on Nei's pairwise genetic distances. Subspecies designation is after Critchfield (1957). The length of the bar indicates the genetic distance between populations. 108

Figure 3-3: Unrooted neighbour-joining tree from the 19 sampled populations of *Pinus contorta* subsp. *latifolia*, based on Nei's pairwise genetic distances. Approximate geographic location is included in parentheses. The length of the bar indicates the genetic distance between populations. 109

Figure 3-4: Unrooted neighbour-joining tree from the 7 sampled populations of *Pinus contorta* subsp. *contorta* and 1 sampled population of *Pinus contorta* subsp. *bolanderi* (136), based on Nei's pairwise genetic distances . Approximate geographic location is included in parentheses. The length of the bar indicates the genetic distance between populations. 110

Figure 3-5: Unrooted neighbour-joining tree from the four sampled populations of *Pinus contorta* subsp. *murrayana*, based on Nei's pairwise genetic distances. Approximate geographic location is included in parentheses. The length of the bar indicates the genetic distance between populations. 111

Figure 3-6: $\hat{M}$ plotted against d on a log-log scale for pairwise comparisons of all 31 populations and for populations within each of the three subspecies in *Pinus contorta*. Scatterplots are shown for $\hat{M}$ calculated using both G_{ST} and $\hat{\theta}$ as estimators of F_{ST} from the equation $\hat{M} = (1/F_{ST} - 1)/4$. 112

Figure 3-7: Approximate extent of glacial ice cover at 18,000 ybp. (After Dyke and Prest 1987). 113

Figure 3-8: Approximate extent of glacial ice cover at 14,000 ybp. (After Dyke and Prest 1987)........ 114

Figure 4-1: Locations of 31 populations of *Pinus contorta* used in this study. 141

Figure 4-2: Comparison of the mitochondrial *nad1* b/c intron repeat region between *Pinus contorta* and *Pinus ponderosa*. 142

Figure 4-3: One of 8,610 most parsimonious trees based on the combined chloroplast *trnL* intron and *trnL/F* spacer regions. * indicates branches that collapse in a strict consensus tree. 143

Figure 4-4: Majority rule consensus tree of 8,610 most parsimonious trees based on the combined chlorplast *trnL* intron and *trnL/F* spacer regions. Numbers above the nodes indicate the percentage of trees that supported the nodes. 144

Chapter 1

Introduction

Introduction

Genetic diversity is the cornerstone upon which the success of a species rests. It is the primary determinant of variation in morphology and physiology, and is required for species to adapt and change with changing environments; variation is the medium upon which the forces of selection act. Genetic variation at multiple levels, within individuals, within populations and among populations, is important for the maintenance of evolutionary potential (Namkoong 1992). Loss of variation can result in inbreeding depression in individuals and expose populations to potentially harmful stochastic forces, due to a lack of ability to respond to selection.

Although the importance of genetic diversity has long been recognized for the above reasons, the first real evidence of variation at the molecular level (for proteins) was determined relatively recently (Hubby and Lewontin 1966; Lewontin and Hubby 1966; Harris 1966). Analysis of isozymes demonstrated that variation in amino acid sequences exists within individuals and populations. Since the introduction of isozyme analysis, research on the amount of variation that exists within and between populations has been conducted over a broad spectrum of species. The amount of variation and the way it is partitioned is affected by a variety of factors such as geographic range, breeding system, dispersal mechanisms, population size, distribution, and environmental and interspecies interaction. Each of these factors can have a different relative impact on genetic diversity. They are included among a list of forces affecting genetic diversity given by

Hartl and Clark (1989) who assign the greatest importance to selection, mutation, mating system, migration, and random genetic drift.

The current distribution of diversity within and among populations of a species, in concert with knowledge of the species' biology (e.g. mating system, population size, edaphic and climatic requirements), allows us to make inferences about past distributions and relationships with other species. Knowledge of the level and how genetic diversity is partitioned allows us not only to look backwards, but forwards as well. It is fundamental for the successful management and conservation of species (Yang and Yeh 1992). For species such as forest trees with long generation times, current human-induced environmental changes may be occurring faster than the rate at which adaptation can occur (Lande 1988). Understanding past responses of populations to climate change will greatly assist future planning.

Species that are long lived, with outcrossing mating systems and large ranges typically have high levels of variation within populations and small among-population differences (Hamrick *et al.* 1992). Most forest trees fall into this category, including lodgepole pine.

Lodgepole pine

Distribution and ecology

Lodgepole pine (*Pinus contorta* Dougl.) is an important component of forest ecosystems in western North America. Its current range extends from the Yukon, through British Columbia and western Alberta, south to California and southeast along the Rocky Mountains into Idaho, Utah, and Colorado (Figure 1-1). This large distribution over extremely varied habitats can be attributed to its tolerance of a wide range of environmental conditions. It is found at sea level along the Pacific coast and at elevations approaching 3900m in the southern Sierra Nevada (Critchfield 1980). *Pinus contorta* is able to withstand the extremes of moisture stress, growing in bogs as well as on excessively drained soils. This has been attributed in part to a finely tuned control of its internal water balance (Lopushinsky 1975; Rundel *et al*. 1977).

The abundance of stands, combined with valued wood characteristics, and relatively short rotation times has resulted in lodgepole pine becoming a focus of foresters, who are interested primarily in its utilization, silviculture and tree improvement strategies (e.g. Dahms 1975; Illingworth 1975). In Alberta its economic importance ranks second after white spruce (Dhir and Barnhardt 1993), and in British Columbia it is the most intensively harvested and planted tree species (B.C. Ministry of Forests 1992).

Pinus contorta is also of great interest to others interested in the biology, genetics and history of forest trees (e.g. Crossley 1955; Wheeler and Guries 1982b; Yang and Yeh 1995; Critchfield 1985). Critchfield (1980) described lodgepole pine as an aggressive pioneering species. It is a shade intolerant early successional species that exhibits rapid early growth, early seed production, and small seed size. In much of its range lodgepole pine has a fire origin, producing serotinous cones that accumulate over time, and provide a large reserve of seeds that are released during fire events. This is contributes to its wide distribution in areas where fire is a dominant natural disturbance.

Cone serotiny, however, is not ubiquitous in lodgepole pine. Populations along the coast of the Pacific and in the Sierra Nevada have primarily open cones. The frequency of stand fire history is assumed to be the major factor that influences the frequency of cone serotiny, although studies examining this are few. One such study by Hoffman and Alexander (1976) showed that in areas where lodgepole pine had become a long-term climax species, cone serotiny was low, allowing seed to germinate and establish in forest gaps. In other areas where fire was an important component of the successional system, cone serotiny was high.

Pinus contorta is a wind-pollinated outcrossing conifer (Critchfield 1980). Like other conifers, the chloroplasts of lodgepole pine are predominately paternally inherited, (Neale *et al.* 1986; Szmidt *et al.* 1987; Wagner *et al.* 1987; Neale and Sederoff 1988) being transmitted through pollen. This mode of inheritance is contrasted with the maternal inheritance of mitochondrial genes through seeds. The two modes of inheritance have been utilized to investigate the relationship and level of introgression of lodgepole pine with jack pine (*Pinus banksiana*). Dong and Wagner (1994) showed that both chloroplast and mitochondrial variants clearly separate the two species, and that due to differential dispersal ability between pollen and seeds, the spread of mitochondrial variants was much greater than the spread of chloroplast variants within species. The maternal inheritance of mitochondria however, is not absolute. Some leakage through pollen does appear to occur, although at a low frequency (Wagner *et al.* 1991).

Systematics

Pinus contorta was placed in a group within the Diploxylon (hard pines) along with *Pinus banksiana*, *Pinus virginiana*, and *Pinus clausa* by Duffield (1952). This group was renamed subsection Contortae by Little and Critchfield (1969) in their revision of the genus, however the members of the subsection were unchanged. A phylogenetic analysis of the subsection based on allozymes confirmed these relationships (Wheeler *et al*. 1983).

Critchfield (1957, 1980) subdivided *Pinus contorta* into four subspecies that are now generally recognized (Figure 1-1). *Pinus contorta* subsp. *contorta* is the coastal variety, restricted to areas proximal to the Pacific Ocean. *Pinus contorta* subsp. *murrayana* ranges from the Cascade Mountains of Oregon south to the Sierra Nevada. *Pinus contorta* subsp. *latifolia* is the most widespread of the subspecies, occupying the Rocky Mountain range and interior British Columbia. *Pinus contorta* subsp. *bolanderi* is restricted to a small geographic area on the coast of California near Mendocino Bay.

Historical reconstruction

The strong differences in morphology between the subspecies of lodgepole pine led Critchfield (1985) to suggest that they have been geographically and genetically isolated for millennia and that glacial events had little impact on the distribution of populations south of the limits of Cordilleran ice. The majority of the current range of lodgepole pine was covered by ice during the last glaciation period, the Wisconsin. As the ice retreated, opportunities for colonization of newly exposed land existed for species in adjacent unglaciated areas.

There are essentially two different methods available for reconstruction of the historical migration of a species. First, evidence can be inferred from the current distribution of the species, based on variations in morphology and the geographic arrangement of alleles. Alternatively, direct evidence from the fossil record can be used to reconstruct vegetation history, providing dates at which particular species arrived in an area. These two methods are not always in agreement.

The early Pleistocene fossil record of lodgepole pine is quite sparse, likely a result of repeated glacial events. Most of the information that does exist refers primarily to periods of time dating back no further than the late Wisconsin. As a result little can be inferred regarding the distribution of lodgepole pine prior to this period.

Subspecies *bolanderi*

The smallest group of the four subspecies is *Pinus contorta* subsp. *bolanderi*. Few investigations have studied this taxon in detail. Critchfield (1985) has proposed that this population has existed in place, throughout glacial events, separated from nearby populations of subsp. *contorta* by the edaphic conditions of the area. Although one might expect that pollen migration from nearby sources of subsp. *contorta* would dilute the integrity of the subspecies over time, Wheeler and Guries (1982a) used isozyme markers to show that a population from subsp. *bolanderi* was genetically distinct from populations belonging to the other subspecies of *Pinus contorta*. In their study, subsp. *bolanderi* came out as a solitary basal group on a phenogram based on Nei's (1972) measure of genetic distance. A more recent study by Aitken and Libby (1994) focused more intensely on subsp. *bolanderi* and its relationship with subsp. *contorta*. These authors found that despite the substantial morphological and physiological differences between the two subspecies, allozyme differentiation was quite weak. This indicates that gene flow may be occurring between the two subspecies or that the time since divergence has been insufficient to allow significant differentiation at allozyme loci. Pollen profiles taken nearby favour the second explanation. Pollen data from Cape Mendocino (120 km. north of *bolanderi* populations) and other locations indicate the probable arrival of *Pinus contorta* at ~3000 years ago during a southward expansion Heusser (1960). This suggests that subsp. *bolanderi* has evolved *in situ* since that time.

Subspecies *contorta*

Pinus contorta subsp. *contorta* is a very hardy subspecies of lodgepole pine. It grows in extreme coastal habitats, occupying niches that are generally unacceptable for potential competitors. In the northern part of its range (Vancouver Island and northward) it exists primarily in bogs and muskeg. In the southern part (south from Vancouver Island) it occupies cliffs and sandy sites. This north / south delineation in habitat was also detected in the chemistry of the trees by Forrest (1980, 1981), who used monoterpenes to evaluate the biogeography of the species. Wheeler and Guries (1982b) however were not able to detect this difference in north / south ecotypes using allozyme analysis. This difference in ecotypes may be due to some selective pressure related to the colder northern environment, or the extreme edaphic conditions in which this subspecies grows. Allozymes, which are generally considered to be selectively neutral, may not reflect this difference unless they were linked to some important trait under selection. Another possible explanation is that the sampling of populations in Wheeler and Guries' (1982b) study was insufficient to detect this difference. Their southernmost sampled population was near the interface of the two ecotypes.

The southern ecotype of subsp. *contorta* most probably originates from populations that survived glaciation south of the ice front (Wheeler and Guries 1982b, Critchfield 1985, Peteet 1991). The origins of the northern ecotype however, are less clear. Hopkins (1972) concluded, based on palynological evidence, that subsp. *contorta* populations south of the ice front quickly expanded northward as the glaciers retreated. The genetic and chemical evidence of Wheeler and Guries (1982b) and Forrest (1980), as well as more recent palynological research, throw some doubt on this hypothesis.

A rapid expansion of populations northward ought to leave detectable genetic characteristics. Founder effects resulting from expansion from a single refugium south of the ice would mean that northern populations should have a close genetic affinity to southern ones, be relatively homogeneous, and have only a small number of rare alleles. Given that the time elapsed since the last glacial retreat is only ~12 000 - 13 000 years B.P., the opportunity for population differentiation would be small. In contrast to these expectations, Wheeler and Guries (1982b) found a large number of rare alleles in northern populations, which were also genetically quite distinct from each other.

The hypothesis of migration from the south is also weakened when the fossil record is considered. Macro-fossil evidence from basal sediments shows that lodgepole pine was present near the Yukon / British Columbia / Alaska border at 10 500 years B.P. (Peteet 1991). This implies that lodgepole pine was present even before the beginning of formation of bogs or peat in the area. Either pine has migrated at an anomalously high rate (>1 km/ year) from populations to the south or subsp. *contorta* survived glaciation in refugia along the coast. Evidence for coastal refugia has been presented by Coulter *et al.* (1965) and Heusser (1960), who examined species distributions in the area.

The fossil evidence showing the presence of lodgepole pine at the northern edge of its current range close to the time of glacial retreat, taken together with the genetic evidence of a separate and distinguishable northern ecotype suggests that northern refugia existed from which current populations of subsp. *contorta* originate. It is unclear whether this northern ecotype migrated southwards to meet northward moving populations from the southern refugium. If this were the case, an increase in the number of rare alleles at this junction of the two ecotypes would support this hypothesis. In Wheeler and Guries (1982b) the highest number of rare alleles in subsp. *contorta* is reported in the Queen Charlotte and North Vancouver Island populations, which is an approximate mid point between putative northern and southern refugia.

Subspecies *murrayana*

The Cascades - Sierra Nevada populations of lodgepole pine (subsp. *murrayana*) have received little detailed attention in terms of their origins. Studies of Quaternary vegetation show the presence of lodgepole pine in the area of the current distribution of subsp. *murrayana* in the early Holocene (Heusser 1983; Barnosky 1984). Although these authors were unable to distinguish the subspecies, other work indicates that it is most likely that of subsp. *murrayana*. The presence of macrofossils near Portland Oregon that date from the late Tertiary period, provides convincing evidence that lodgepole pine was in the area before the Pleistocene (Critchfield 1980, based on personal communication with Wolfe 1971). Axelrod and Ting (1961) working in the Sierra Nevada also identified pollen from the early Pleistocene as being subsp. *murrayana*.

A recent study by Anderson (1996) attempted to reconstruct the biogeography of subsp. *murrayana,* in the Sierra Nevada during the late Wisconsin deglaciation. Although this area was unglaciated, the influence of global glaciation, and resulting climate change would have affected the distribution of species. Macrofossil evidence showed that this pine was present at least 500m lower in elevation during this period than in its current range. This author suggested that between 9000 and 6750 years B.P. climatic warming caused drying of soils in the lower elevational areas, resulting in a movement of populations to higher (less dry) elevations. This elevational migration is in contrast with subsp. *latifolia*, which had both an elevational and a large latitudinal migration.

It is interesting that although both subsp. *murrayana* and *latifolia* were present near the ice edge, and were potentially able to advance during ice retreat, only subsp. *latifolia* was able to expand northward after deglaciation. This may be in part due to differences in cone serotiny between these two subspecies. Subsp. *latifolia* has primarily closed cones, whereas cone serotiny is non existent in the subsp. *murrayana* in the Sierra Nevada and rare in the southern Cascades (Critchfield 1957; Mowat 1960). As weather became warmer and more arid during the Holocene, fire frequency increased (Barnosky 1984). This increase in forest fires may have allowed the more fire dependent subsp. *latifolia* to expand quickly into new areas (Cwynar 1987) while subsp. *murrayana* moved primarily upwards in elevation, pursuing the cooler, moister conditions.

Subspecies *latifolia*

The largest amount of information on the post-glacial history of lodgepole pine exists for subsp. *latifolia*. This is perhaps due to the larger geographical area it currently covers. Fossil pollen evidence throughout the region shows that the Rocky Mountain - Intermountain range was dominated by *Picea* at approximately 10,000 years B.P. (MacDonald 1987a, b; Cwynar 1988). *Pinus contorta* subsp. *latifolia* extended its range northward over the course of the Holocene (MacDonald and Cwynar 1991) and appears to still be expanding in the north (MacDonald and Cwynar 1985; Critchfield 1985)

A model of expansion was presented by MacDonald and Cwynar (1991). They collected fossil pollen data that is consistent with a pattern of expansion of subsp. *latifolia* populations via successive establishment of small populations that then increased to current densities over periods of 1,000 to 4,000 years.

It is well accepted that a region in the west-central Yukon territory was unglaciated during the Wisconsin (e.g., Hughes *et al.* 1968; Dyke and Prest 1987). Because this area today supports lodgepole pine populations, Hulten (1937) suggested that lodgepole pine survived glaciation in this northern refugium. Matthews (1970) and Lichti-Federovitch (1974) showed using palynological data that lodgepole pine was present in the Yukon during the last interglacial period of the Pleistocene. Other researchers have supported the hypothesis that populations survived the Wisconsin glaciation in a Yukon refugium. In provenance trials (using populations from the northern part of the range of subsp. *latifolia*), Hagner (1970) detected consistent differences in Yukon populations compared with others, despite that no clear pattern was detected. Some indirect measures of cold hardiness indicated that the Yukon populations were similar to populations 1,000 km to the south, while other measures indicated that Yukon populations have a greater cold hardiness than adjacent northern populations.

A number of genetic studies support the hypothesis of a Yukon refugium. Yeh and Layton (1979), in a comparison of central and marginal populations of subsp. *latifolia*, demonstrated the distinctiveness of Yukon populations. The Yukon populations in their study lacked one allele present in all other populations and also possessed one allele not present anywhere else. Another study by von Rudloff and Nyland (1979) also demonstrated the unique character of Yukon populations compared to southern ones, based on resin monoterpene composition. Wheeler and Guries (1982b), using clustering techniques, showed that populations from the Yukon and northern British Columbia formed a cohesive group. As well, a number of rare alleles in Yukon populations were absent in other populations of subsp. *latifolia*. The clustering of populations in this study also indicates the potential for two routes of northward migration from southern refugia. One route follows the slopes of the Rocky Mountains, while the other is through the central interior of British Columbia. A central interior migration route is also supported by palynological evidence presented by Mack et al. (1978).

MacDonald and Cwynar (1985) contested the hypothesis of persistence of lodgepole pine during the Wisconsin in a Yukon glacial refugium. Using fossil pollen, they show a clear south to north geographic trend in arrival dates of lodgepole pine. The pollen record indicates that lodgepole pine was present in southwestern Alberta ~12,200 years B.P. It expanded its range northward, reaching Spring Lake (north of Jasper) by 10, 800 years B.P. and moved north of 60° latitude by 5,600 years B.P. The trend continues west and north to a population in the Yukon that appears to have been established no earlier than 430 years B.P. Based on the fact that lodgepole pine is a prolific pollen producer, they view the fossil pollen evidence as providing conclusive evidence that lodgepole pine did not persist in northern refugia.

Lodgepole pine is a prolific pollen producer (approximately 500 million to 20 billion pollen grains per tree (Critchfield 1985)), and the small pollen size potentially permits very long distance transport. Because of this, MacDonald and Cwynar (1985) used a value of 15% (of total pollen composition per sample) as a cut-off point to indicate the presence of lodgepole pine. As noted by Peteet (1991), this value is based on relationships developed between the present distribution of lodgepole pine in the western interior and modern pollen deposition (see MacDonald and Ritchie 1986). Peteet (1991) however has shown conclusively (based on macrofossils) that *Pinus* was present near Yakutat, Alaska even though pollen percentages in the same strata were less than 2%. She also notes that the pine in the area of her study currently produce minimal amounts of pollen. If this is a result of edaphic or climatic constraints, a similar situation may exist in other populations at environmental extremes.

Another possible origin of Yukon populations of lodgepole pine is from north coastal sources. Wheeler and Guries (1982b) note that many of the rare alleles found in Yukon populations are also shared by subsp. *contorta* populations along the northern coast. This hypothesis was investigated by Spear and Cwynar (1997). Using primarily fossil pollen evidence from the White Pass in northern British Columbia (a major pass linking the interior of B.C. and Yukon with coastal Alaska) they established that trees and shrubs were not able to use the pass as a migration route from coastal to interior areas. These authors estimated that a population of lodgepole pine at Waterdevil Lake, just north of the pass, arrived at 3,200 years B.P. The nearest population (90 km to the northeast) appears to have arrived at a later date ~ 2,500 years B.P. Without evidence that lodgepole pine migrated through the pass, these authors were unable to explain the origin of pine at Waterdevil Lake, but suggest migration of coastal pine occurred up an alternate valley.

Evolutionary forces

Population genetics is concerned with describing the amount and distribution of variation in populations, along with the various forces that shape these parameters. Foremost in the description of populations is the degree of heterozygosity and homozygosity (or level of fixation), and the relationship of these parameters to what is expected given certain conditions. The fixation index, as well as other parameters, such as genetic distance, is determined by the relative influence of forces such as random drift, mutation, migration, breeding system and selection.

For some taxa, such as *Pinus contorta* subsp. *latifolia*, post glacial colonization proceeded in a stepwise pattern of long distance dispersal and subsequent population growth (MacDonald and Cwynar 1985). Each long distance dispersal event would contain seeds sampled from the source population. Sampling in this manner is analogous to founder effects, or to the effect of genetic drift on populations (Yeh 2000).

The effects of random genetic drift are important in populations of small size. In any given generation, only a small fraction of the available gametes are combined in offspring to form the next generation. This sampling of the gametes will result in chance differences in allele frequencies between the sample and the pool of gametes. For a locus with two alleles (*A* and *a*) the probability of the sample containing *i* alleles of type *A* is given by the binomial probability:

$$\textit{Probability of (i) alleles in the sample} = \frac{(2N)!}{i!(2N-i)!}p^i q^{2N-i}$$

where *N* is the population size, p is the allele frequency of *A*, and $q = 1-p$. Once one or the other allele becomes fixed in the population (i.e. $A = 0$ or 1) the alternate allele is lost. The rate of fixation of alleles depends not only on the size of the population, but also the initial frequency of the allele. In fact the probability of a particular neutral allele becoming fixed at some point is equal to its frequency in the population (Hartl and Clark 1989). As alleles become fixed, heterozygosity naturally decreases. Expected heterozygosity at some generation (t) in the future is therefore a function of the initial heterozygosity (H_0) and the size of the population (N):

$$H_t = H_0\left(1-\frac{1}{2N}\right)^t$$

Founding populations arising from long distance dispersal will therefore likely have different allele frequencies from the source population (due to initial sampling) and be prone to degradation in allelic diversity and heterozygosity (from continuing random drift). According to Nei *et al.* (1975) however, reduced heterozygosity as a result of founding events is only expected if the initial number of founding individuals is small, and the population increases in size only slowly.

Founder effects have been detected in a number of species. Ledig (2000) showed that *Pinus coulteri* (Coulter pine) exhibits a reduction in allelic diversity that correlates with latitude. Ledig surmised that the most plausible explanation for this observation was a cascading founder effect associated with long distance dispersal during northward postglacial migration. Other examples of reduced diversity along a north-south cline that may be a result of founder effects are seen in *Helonias bullata* (Hamrick and Godt 1996), *Casuarina cunninghamiana* (Moran *et al*. 1989), and *Pinus jeffreyi* (Furnier and Adams 1986).

As mentioned, variation is a key requirement in evolutionary process. The ultimate source of variation is mutation. The term mutation however, is often broadly applied to a number of different processes. On a large scale, this includes chromosomal rearrangements such as inversions or translocations, and changes in ploidy number. At the protein level, these include amino acid substitutions, and insertions and deletions. The simplest process that the term is applied to is that of a point mutation in which a single nucleotide of a DNA sequence is changed. Such an event may or may not have an effect on the phenotype of the organism depending on whether the mutation occurs in a region of DNA that encodes a gene product or is involved in control of gene processing, and whether the mutation results in a synonymous or non-synonymous change. When a mutation rate is estimated, it is typically point mutations that are being referred to. For most species studied to date, mutation rates are of the order 10^{-8} to 10^{-11} substitutions per nucleotide site per year (Futuyma 1998). The rate of mutation however is not constant across the genome; it can vary from gene to gene, and even regions within genes can have different rates. The mutation rate μ is related to the fixation index F by the equation $F = \frac{1}{4N\mu + 1}$. Thus, as the mutation rate increases, the probability of fixation at a particular locus (due to drift) decreases.

The relationship between migration (gene flow) and the fixation index is analogous to that of mutation and F. At migration rate m, $F = \frac{1}{4Nm + 1}$. As migration increases, the probability of a particular locus becoming fixed decreases. Migration between populations can have a much greater moderating effect on the fixation index, because migration rates are typically much greater than mutation rates. In fact, a small amount of migration can prevent substantial divergence between populations due to the effects of random drift (Hartl and Clark 1989). When there are large amounts of migration between all populations of a species, the effect of drift can be reduced to the point where it is essentially acting as if there were no population structure. The amount of migration that occurs between populations is often a function of the distance between populations. This concept is incorporated into the 'isolation by distance' model of gene flow (Wright 1943) and the 'stepping stone model' (Kimura and Weiss 1964). Both models predict that differentiation between populations will increase, the further apart (geographically) they are. The main difference between these two models is that populations in the 'stepping stone' model are discrete and separate from one another. In the 'isolation by distance' model, populations are continuously distributed such as in many forest tree species. Migration, or gene flow, can refer to individuals such as movement of animals between populations, or more generally as movement of gametophytes. The amount of gene flow between populations is affected not only by the physical structure of populations, but also by dispersal ability. For example, in wind pollinated plants or marine organisms with easily dispersed larval stages, gene flow can be quite high.

It is important to reiterate that the above equations describing the relationship between mutation and migration with the fixation index hold true for 'ideal' populations at equilibrium between these forces and drift. For many species, populations may not be at equilibrium, for a variety of factors.

The degree of fixation or homozygosity of an individual is also closely linked with the breeding system of the species in question. Monoecious species can be self-fertilizing, as is common in grasses or primarily outcrossing, as is seen in most forest trees. For outcrossing species the amount of selfing that occurs is usually small due to temporal differences in the development of male and female structures, physical separation of male and female structures, and self-incompatibility mechanisms. With repeated selfing, heterozygosity is halved every generation (Hartl and Clark 1989). After even a few generations, the level of fixation approaches 1. The effect of partial selfing can also quickly result in reduced heterozygosity and lead to inbreeding depression in species that are predominantly outcrossing. If substantial inbreeding is occurring, this naturally implies that no (or very little) migration between populations is happening. In such a situation, the effects of random drift can be enhanced, leading to increased among-population differentiation.

The importance of selection as a process in evolution has led to the development of a number of theories and models to describe and quantify selection, or the fitness of certain genotypes and phenotypes. The effect of selection at a particular locus can be dramatic; an advantageous allele can quickly become increasingly widespread over successive generations. In finite populations, random drift can have the effect of accelerating the rate to fixation of such an allele, or retarding such an outcome. The size of the population (N) and the selection intensity (s) of the allele in question affect the relative probabilities. When populations are large in size the effects of drift are small, so that an allele with a modest s can become widespread. At small population sizes, even a very advantageous allele may not become fixed in a population, due to drift. When the product $Ns > 1$, fixation of the advantageous allele is likely, whereas at values for $Ns < 0.1$, the effect of selection is negligible and the chance of fixation is the same as that under a model for a neutral allele (i.e., a function of initial allele frequency) (Hartl and Clark 1989). Because selection for an allele can be frequency dependent, associated with heterozygosity (heterozygotes superiority or inferiority), or fecundity, and also affected by drift, determining the relative importance of the various forces involved can be difficult. When determining population relationships, it is therefore advantageous to examine loci that are not experiencing strong selection pressure.

RAPD markers

Much of the available information concerning the genetic variability of lodgepole pine is from allozyme studies (e.g. Wheeler and Guries 1982a; Wheeler and Guries 1982b; Yang and Yeh 1995). In a simulation study of the effects of balanced polymorphisms, background selection and local selection on genetic diversity in subdivided populations, Charlesworth *et al.* (1997) suggested that F_{ST} for unlinked loci would be much lower than that for loci linked to loci under local selection. For example, if locus A was linked to locus B, and locus B was experiencing divergent selection across two populations, then an estimate of F_{ST} based on locus A might be inflated. Hence, the use of F_{ST} to measure the between-population differentiation could be ambiguous when comparisons were made between genome regions with different rates of crossing over. Even in outcrossing populations, local selection on certain loci in the genome could produce large differences in allele frequency at other loci when gene flow was restricted. Charlesworth *et al.* (1997) proposed that genetic markers reflecting neutral differences be used to measure the between-population diversity because these relate most directly to the coalescence times. They further suggested that data on several loci should be used.

Random amplified polymorphic DNA (RAPD) markers are well suited for population genetic analysis in that they are believed to be primarily neutral and are highly variable. Each random primer typically yields several variable RAPDs; and since many random primers are available, a large number of RAPDs are available for analysis. RAPDs have been shown to be inherited in a biparental dominant Mendelian manner (Carlson *et al.* 1991; Roy *et al.* 1992; Heun and Helentjaris 1993). Their utility as markers for the genetic characterization of populations has been well established for a variety of organisms (Chalmers *et al.* 1992; Huff *et al.* 1993; Yeh *et al.* 1995; Caccone *et al.* 1997; Hogbin *et al.* 1997; Esquibet *et al.* 1998; Nkongolo *et al.* 2002).

Sequence data

Although the application of molecular markers in population genetics has had a fundamental importance in evaluating levels of polymorphism and heterozygosity, and has been useful for evaluating models and expectations, DNA sequence data is the definitive source of genomic information. Advances in technology and widespread interest in have resulted in an exponential increase in the amount of DNA sequence that has been resolved. To date, complete genome sequences are available from over 100 microbial species, as well as chloroplast and mitochondrial genomes of a number of plant and animal species. The large size of the nuclear genome of eukaryotes has resulted in complete sequencing of only a handful of species (such as *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Saccaromyces cerevisiae*. *Mus musculus*, and *Homo sapiens*), but detailed genomic maps for many species are available. Concurrent advances in computing power have facilitated the development of programs to deal with the large data sets and perform complex analyses.

DNA sequence data obtained from related species has been very useful for resolving phylogenies at a number of levels. For example, analysis of ribosomal DNA sequence has shown that Gnetales, long considered to be a sister group to the angiosperms, are more closely related to the gymnosperms, and perhaps a sister group to the Pinaceae (Liston *et al.* 1996; Chaw *et al.* 1997; Chaw *et al.* 2000). Within families and genera, comparision of sequence data have further resolved weak phylogenies derived from morphological characters and given insight into the origins of particular species (Olsen and Schaal 1999). Restriction fragment analyses of sequence data have also revealed variation within populations of species (Strauss *et al*. 1993; Wu *et al*. 1998; Sinclair *et al*. 1999)

Objectives

By studying species such as lodgepole pine, insight into the genetic processes affecting parameteres such as the effects of gene flow, genetic drift, diversity can be gained. Lodgepole pine in particular is an interesting study species because the processes influencing population genetic structure can be evaluated in each of its subspecies. This provides a comparison of different factors between very closely related groups.

The objective of my thesis research is to investigate the population genetics of lodgepole pine with an emphasis on the relationship between the subspecies and the effects of post-glacial migration on various population genetic parameters and patterns. In Chapter 2, a set of RAPD markers are used to compare genetic diversity between marginal, intermediate and central populations of subsp. *latifolia*. Estimates of population differentiation as well as an analysis of multilocus associations are used to reveal additional population structure. For Chapter 3, an expanded RAPD data set including all four subspecies is generated to evaluate the hypothesis of isolation by distance, and the potential for northern refugia of lodgepole pine during the Wisconsin glaciation. In Chapter 4, sequence data from chloroplast and mitochondrial regions are obtained with the objective of assessing the relationships among the subspecies and further exploring postglacial movement of the species. Chapter 5 provides an overall overview of the results with suggestions for avenues of further research.

References

Aitken, S.N. and W.J. Libby. 1994. Evolution of the pygmy-forest edaphic subspecies of *Pinus contorta* across and ecological staircase. Evolution **48**: 1009-1019.

Anderson, R.S. 1996. Postglacial biogeography of Sierra lodgepole pine (*Pinus contorta* var. *murrayana*) in California. Ecoscience **3**: 343-351.

Axelrod, D.I. and W.S. Ting. 1961. Early Pleistocene floras from the Chagoopa surface, southern Sierra Nevada. Univ. Calif. Publ. Geol. Sci. **39**: 119-194.

Barnosky, C.W. 1984. Late Pleistocene and early Holocene environmental history of southwestern Washington State, U.S.A. Can. J. Earth Sci. **21**: 619-629.

British Columbia Ministry of Forests. 1992. Annual report of the Ministry of Forests, 1991-1992. British Columbia Ministry of Forests, Victoria, B.C.

Caccone, A., G. Allegrucci, C. Fortunato and V. Sbordoni. 1997. Genetic differentiation within the European sea bass (*D. labrax*) as revealed by RAPD-PCR assays. J. Hered. **88**: 316-324.

Calder, J. and R. Taylor. 1968. Flora of the Queen Charlotte Islands. Can. Dep. Agr. Res. Branch Monograph No. 4, part 1.

Carlson, J.E., L.K. Tulsieram, J.C. Glaubitz, V.W.K Luk, C. Kauffeldt and R. Rutledge. 1991. Segregation of random amplified DNA markers in F_1 progeny of conifers. Theor. Appl. Genet. **83**: 194-200.

Chalmers, K.J., R. Waugh, J.I. Sprent, A.J. Simons and W. Powell. 1992. Detection of genetic variation between and within populations of *Gliricidia sepium* and *G. maculata* using RAPD markers. Heredity **69**: 465-472.

Charlesworth, B., M. Nordborg and D. Charlesworth. 1997. The effect of local selection, balanced polymorphism, and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res., Camb. **97**: 175-181.

Chaw, S.-M., A. Zharkikh, H.-M. Sung, T.-C. Lau and W.-H. Li. 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. Mol. Biol. Evol. **14**:56-68.

Chaw, S.-M., C.L. Parkinson, Y. Cheng, T.M. Vincent and J.D. Palmer. 2000. Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. Proc. Nat. Acad. Sci. **97**: 4086-4091.

Coulter, H.W.D., D.M. Hopkins, T.N.V. Karlstrom, T.L. Pewie, C. Wahrhaftig and J.R. Williams. 1965. Map showing extent of glaciation in Alaska. U.S. Geol. Surv. Misc. Invest. Map No. I-415.

Critchfield, W.B. 1957. Geographic variation in *Pinus contorta*. Maria Moors Cabot Foundation. Publ. 3. Harvard University Press.

Critchfield, W.B. 1980. Genetics of Lodgepole Pine. U.S. For. Serv. Res. Pap. WO-37.

Critchfield, W.B. 1985. The late Quaternary history of lodgepole and jack pines. Can J. For. Res. **15**: 749-772.

Crossley, D.I. 1955. The production and dispersal of lodgepole pine seed. Can. For. Br., For. Res. Div. Tech. Note 25.

Cwynar, L.C. 1987. Fire and the forest history of the North Cascade range. Ecology **68**: 791-802.

Cwynar, L.C. 1988. Late Quaternary vegetation history of Kettlehole Pond, southwestern Yukon. Can. J. For. Res. **18**: 1270-1279.

Dahms, W.G. 1975. Gross yield of central Oregon lodgepole pine. *In:* Management of lodgepole pine ecosystems: symposium proceedings, vol.1. Coop. Ext. Serv., Wash. State Univ., Pullman. pp. 208-232.

Dhir, N.K. and L.K. Barnhardt. 1993. Genetic improvement of lodgepole pine in Alberta. *In*: *Pinus contorta*: From untamed forest to domesticated crop. Edited by D. Lindgren. Swedish University of Agricultural Science, Department of Forest Genetics and Plant Physiology, Sweden. Rep 11. pp. 279-288.

Dong, J. and D.B. Wagner. 1994. Paternally inherited chloroplast polymorphism in *Pinus*: estimation of diversity and population subdivision, and tests of disequilibrium with a maternally inherited mitochondrial polymorphism. Genetics **136**: 1187-1194.

Duffield, J.W. 1952. Relationships and species hybridizations in the genus *Pinus*. Z. Forstgenet. Forstplanzenzücht. [Silvae Genet.] **1**: 93-100.

Dyke, A.S. and V.K. Prest. 1987. Late Wisconsinan history of the Laurentide ice sheet. Geog. Phys. Quater. **41**: 327-364.

Esquibet, M., S. Bekal, P. Castagnone-Sereno, J.P. Gauthier, R. Rivoal and G. Caubel. 1998. Differentiation of normal and giant *Vicia faba* populations of the stem nematode *Ditylenchus dipsaci*: agreement between RAPD and phenotypic characteristics. Heredity **81**: 291-298.

Forrest, G.I. 1980. Geographic variation in the monoterpenes of *Pinus contorta* oleoresin. Biochem. Syst. Ecol. **8**: 343-359.

Forrest, G.I. 1981. Geographical variation in oleoresin monoterpene composition of *Pinus contorta* from natural stands and planted seed collections. Biochem. Syst. Ecol. **9**: 97-103.

Furnier, G.R. and W.T. Adams. 1986. Geographic patterns of allozyme variation in Jeffrey pine. Am. J. Bot. **73**: 1009-1015.

Futuyma, D.J. 1998. Evolutionary Biology. Sinauer Associates, Inc., Sunderland, Massachusetts.

Hagner, M. 1970. A genecological investigation of the annual rhythm of *Pinus contorta* Dougl. and a comparison with *Pinus silvestris* L. Stud. For. Suecica **81**. Roy. Coll. For. Stockholm.

Hamrick, J.L. and M.J.W. Godt. 1990. Allozyme diversity in plant species. *In*: Plant population genetics, breeding, and genetic resources. Edited by A.H.D. Brown, M.T. Clegg, A.L. Kahler, and B.S. Weir. Sinauer Associates Inc., Sunderland, Massachusetts. pp. 44-64.

Hartl, D.L. and A.G. Clark. 1989. Principles of population genetics. Sinauer Associates, Inc., Sunderland, Massachusetts.

Heun, M. and Helentjaris, T. 1993. Inheritance of RAPDs in F_1 hybrids of corn. Theor. Appl. Genet. **85**: 961-968.

Heusser, C.J. 1960. Late Pleistocene environments of north Pacific North America. Am. Geogr. Soc., Special Publ. No. 35.

Heusser, C.J. 1983. Vegetation history of the northwestern United States including Alaska. *In*: Late Quaternary environments of the United States. vol 1. Edited by S.C. Porter. University of Minnesota Press. pp. 239-258.

Hoffman, G.R. and R.R. Alexander. 1976. Forest vegetation of the Bighorn Mountains: a habitat type classification. USDA For. Serv. Res. Pap. RM-170.

Hogbin, P.M., Ayre, D.J., and Whelan, R.J. 1998. Genetic variation and reproductive success of road verge populations of the rare shrub *Grevilla barklyana* (Proteaceae). Heredity **80**: 180-186.

Hopkins, D.M. 1972. The paleogeography and climatic history of Beringia during late Cenozoic time. Inter-nord. **12**: 121-150.

Hughes, O.L., R.B. Campbell, J.E. Muller and J.O. Wheeler. 1968. Glacial limits and flow patterns, Yukon Territory, south of 65 degrees north latitude. Geol. Surv. Can. Geol. Surv. Pap. No. 68-34.

Hulten, E. 1937. Outline of the history of Arctic and boreal biota during the Quaternary period. Bokforlags Aktiebolaget Thule, Stockholm.

Illingworth, K. 1975. Lodgepole pine provenance research and breeding in British Columbia. *In*: Management of lodgepole pine ecosystems: symposium proceedings, vol.1. Coop. Ext. Serv., Wash. State Univ., Pullman. pp. 47-67.

Kimura, M. and G.H Weiss. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics **49**: 561-576.

Lande, R. 1988. Genetics and demography in biological conservation. Science **241**: 1455-1460.

Ledig, F.T. 2000. Founder effects and the genetic structure of coulter pine. J. Hered. **91**: 307-315.

Li, Changxi. 1998. Genetic dissection of resistance to western gall rust (*Endocronartium harknessii* (J. P. Moore) Y. Hiratsuka) in lodgepole pine (*Pinus contorta* var.

latifolia Engelm.). Ph.D. Thesis. University of Alberta. Edmonton, Alberta, Canada.

Lichti-Federovich, S. 1974. Palynology of two sections of late Quaternary sediments from the Porcupine River, Yukon Territory. Can. Geol. Surv. Pap. 74-23.

Liston, A., W.A. Robinson, J.M. Oliphant and E.R. Alvarez-Buylla. 1996. Length variation in the nuclear ribosomal DNA internal transcribed spacer region of non-flowering seed plants. Syst. Bot. **21**: 109-120.

Little, E.L., and W.B. Critchfield. 1969. Subdivisions of the genus *Pinus* (pines). U.S. Dep. Agric., Misc. Publ. 1144.

Lopushinsky, W. 1975. Water relations and photosynthesis in lodgepole pine. *In*: Management of lodgepole pine ecosystems: symposium proceedings, vol.1. Coop. Ext. Serv., Wash. State Univ., Pullman. pp. 135-153.

MacDonald, G.M. 1987a. Postglacial development of the subalpine-boreal transition forest of western Canada. J. Ecol. **75**: 303-320.

MacDonald, G.M. 1987b. Postglacial vegetation history of the Mackenzie River basin. Quater. Res. **28**: 245-262.

MacDonald, G.M. and L.C. Cwynar. 1985. A fossil pollen based reconstruction of the late Quaternary history of lodgepole pine (*Pinus contorta* ssp. *latifolia*) in the western interior of Canada. Can J. For. Res. **6**: 1039-1044.

MacDonald, G.M. and L.C. Cwynar. 1991. Post-glacial population growth rates of *Pinus contorta* ssp. *latifolia* in western Canada. J. Ecol. **79**: 417-429.

MacDonald, G.M. and J.C. Ritchie. 1986. Modern pollen spectra from the western interior of Canada and the interpretation of late Quaternary vegetation development. New Phytol. **102**: 245-268.

Mack, R.N., N.W. Ruther and S. Valastro. 1978. Late Quaternary pollen record from the Sanpoil River valley, Washington, Can. J. Bot. **56**: 1642-1650.

Matthews, J.V. Jr. 1970. Quaternary environmental history of interior Alaska: pollen samples from organic colluvium and peats. Arct. and Alp. Res. **2**: 241-251.

Mowat, E.L. 1960. No serotinous cones on central Oregon lodgepole pine. J. For. **58**: 118-119.

Moran, G.F., J.C. Bell and J.W. Turnbull. 1989. A cline in genetic diversity in river she-oak *Casuarina cunninghamiana*. Aust. J. Bot. **37**: 169-180.

Namkoong, G. 1992. Biodiversity –issues in genetics, forestry and ethics. For. Chron. **68**: 438-443.

Neale, D.B. and R.R. Sederoff. 1988. Inheritance and evolution of conifer organelle genomes. *In*: Genetic manipulation of woody plants. Edited by J.W. Hanover and D.E. Keathley. Plenum Press, New York. pp. 251-264.

Neale, D.B., N.C. Wheeler and R.W. Allard. 1986. Paternal inheritance of chloroplast DNA in Douglas-fir. Can. J. For. Res. **16**: 1152-1154.

Nei, M. 1972. Genetic distance between populations. Am. Nat. **106**: 283-292.

Nei, M., T. Maruyama and R. Chakraborty. 1975. The bottleneck effect and genetic variability in populations. Evolution **29**: 1-10.

Nkongolo, K.K., P. Michael and W.S. Gratton. 2002. Identification and characterization of RAPD markers inferring genetic relationships among Pine species. Genome **45**: 51-58.

Peteet, D.M. 1991. Postglacial migration history of lodgepole pine near Yakutat, Alaska. Can. J. Bot. **69**: 786-796.

Olsen, K.M. and B.A. Schaal. 1999. Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*. Proc. Nat. Acad. Sci. **96**: 5586-5591.

Roy, A., Frascaria, N., Mackay, J., and Bousquet, J. 1992. Segregating random amplified polymorphic DNAs (RAPDs) in *Betula alleghaniensis*. Theor. Appl. Genet. **85**: 173-180.

Rundel, P.W., D.J. Parsons, and D.T. Gordon. 1977. Montane and subalpine vegetation of the Sierra Nevada and Cascade Ranges. *In*: Terrestrial vegetation of California. Edited by M.G. Barbour and J. Major. John Wiley and Sons. pp.559-599.

Sinclair, W.T., J.D. Morman and R.A. Ennos. 1999. The postglacial history of Scots pine (*Pinus sylvestris* L.) in western Europe: evidence from mitochondrial DNA variation. Mol. Ecol. **8**: 83-88.

Strauss, S.H., Y.-P. Hong and V.D. Hipkins. 1993. High levels of population differentiation for mitochondrial DNA haplotypes in *Pinus radiata*, *muricata* and *attenuata*. Theor. Appl. Genet. **86**: 605-611.

Szmidt, A.E., T. Alden and J.-H. Hällgreen. 1987. Paternal inheritance of chloroplast DNA in *Larix*. Plant Mol. Biol. Rep. **9**: 59-64.

Spear, R.W. and L.C. Cwynar. 1997. Late Quaternary vegetation history of White Pass, northern British Columbia, Canada. Arc. and Alp. Res. **29**: 45-52.

Von Rudloff, E. and E. Nyland. 1979. Chemosystematic studies in the genus Pinus. III. The leaf oil terpene composition of lodgepole pine from the Yukon Territory. Can. J. Bot. **57**: 1367-1370.

Wagner, D.B., G.R. Furnier, M.A. Saghai-Maroof, S.M. Williams, B.P. Dancik and R.W. Allard. 1987. Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. Proc. Natl. Acad. Sci. **84**: 2097-2100.

Wheeler, N.C. and W.B. Critchfield. 1985. The distribution and botanical characteristics of lodgepole pine: Biogeographical and management implications. *In*: Lodgepole pine: The species and its management. Edited by D.M. Baumgartner, R.G. Krebill, J.T Arnott and G.F. Weetmen. Coop. Ext. Serv., Wash. State Univ., Pullman. pp. 1-13.

Wheeler, N.C., R.P. Guries and D.M. O'Malley. 1983. Biosystematics of the genus *Pinus*, subsection Contortae. Bioch. Syst. Ecol. **11**: 333-340.

Wheeler, N.C. and R.P. Guries. 1982a. Population structure, genetic diversity, and morphological variation in *Pinus contorta*. Can. J. Bot. **12**: 595-606.

Wheeler, N.C. and R.P. Guries. 1982b. Biogeography of lodgepole pine. Can J. Bot. **60**: 1805-1814.

Wright, S. 1943. Isolation by distance. Genetics **28**: 114-118.

Wu, J., K.V. Krutovskii and S.H. Strauss. 1998. Abundant mitochondrial genome diversity, population differentiation and convergent evolution in pines. Genetics **150**: 1605-1614.

Yang, R.-C. and F.C. Yeh. 1992. Genetic consequences of in situ and ex situ conservation of forest trees. For. Chron. **68**: 720-729.

Yang, R.-C. and F.C. Yeh. 1995. Patterns of gene flow and geographic structure in *Pinus contorta*. Forest Genetics. **2**: 65-75.

Yeh, F.C. 2000. Population genetics. *In*: Forest conservation genetics. Edited by A. Young, D. Boshier and T. Boyle. CABI Publishing, Wallingford, England / CSIRO Publishing, Collingwood, Victoria. pp. 21-37.

Yeh, F.C. and C. Layton. 1979. The organization of genetic variability in central and marginal populations of lodgepole pine *Pinus contorta* ssp. *latifolia*. Can. J. Genet. ytol. **21**: 487-503.

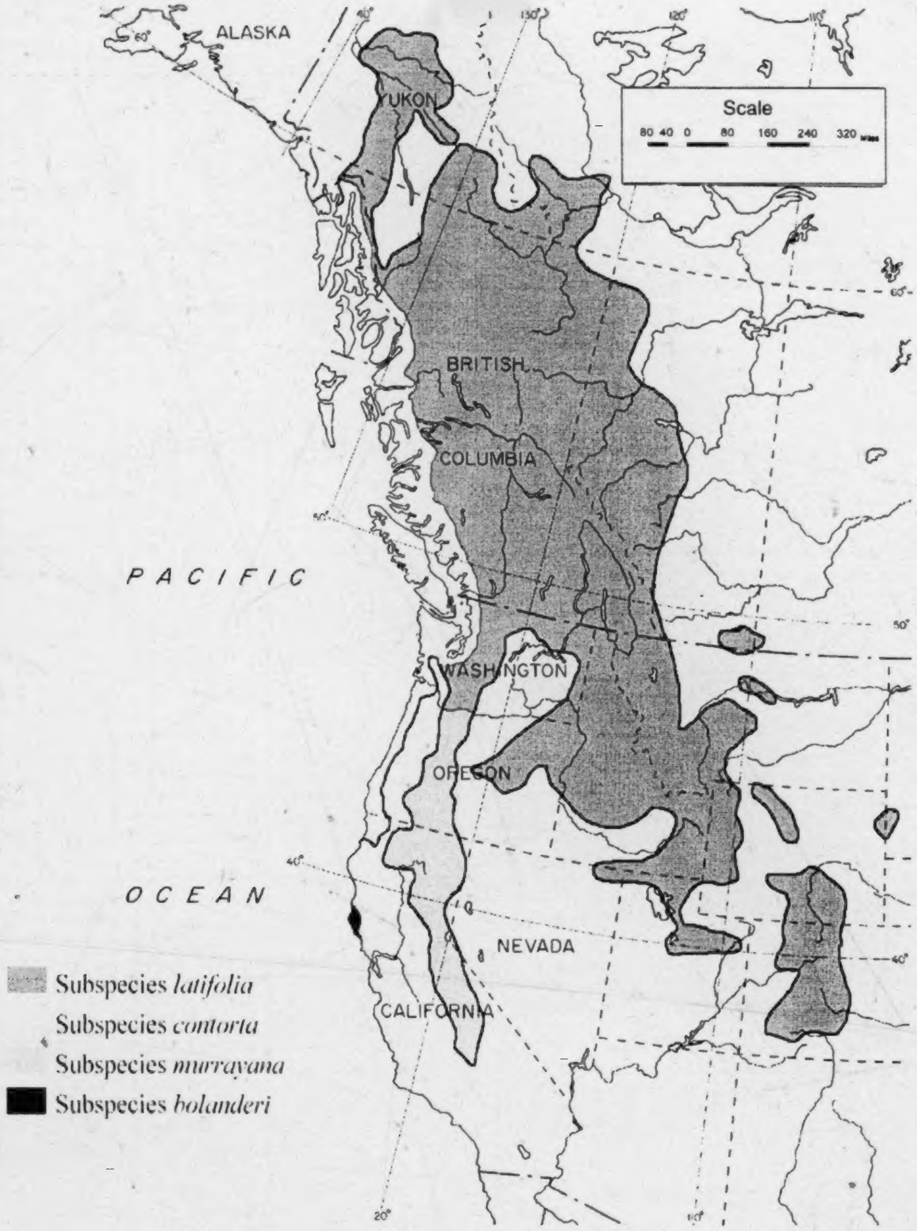
ALASKA
YUKON
BRITISH
COLUMBIA
WASHINGTON
OREGON
NEVADA
CALIFORNIA
PACIFIC
OCEAN
Scale
80 40 0 80 160 240 320 miles
60°
140°
130°
120°
110°
50°
40°
20°
▨ Subspecies *latifolia*
Subspecies *contorta*
Subspecies *murrayana*
■ Subspecies *bolanderi*

Figure 1-1: Approximate distribution of the four subspecies of *Pinus contorta*.

Chapter 2

Random amplified polymorphic DNA diversity of marginal and central populations in *Pinus contorta* subsp. *latifolia* [1]

Introduction

Conifers with broad geographic range generally exhibit large amounts of variation and little population differentiation at allozyme loci (Yeh *et al.* 1994). Extensive gene flow, outcrossing mode, large effective population size, longevity and high fecundity have been cited as determinants of their genetic structure (Loveless and Hamrick 1984; Hamrick and Godt 1990). Disjunct distribution and environmental gradients (Xie *et al.* 1992) and multilocus Wahlund and founder effects (Yang and Yeh 1993) also could influence the genetic structure of conifers. However, little is known about the genetic structure of marginal populations on the edge of the species range. As regards the genetic structure of such populations, epistatic selection, Wahlund and founder effects, chance events from population subdivision, and other evolutionary mechanisms might have a relative importance strikingly different from that reported for central populations.

Reduced genetic variability in marginal populations relative to central ones has been theorized (Mayr 1970) and studied in a variety of organisms. Although initial investigations of this idea showed no clear population trend (Prakash 1973; Arroyo 1975; Dessauer *et al.* 1975; Levin 1977), many subsequent investigations provided support for the concept. This reduced genetic variability in marginal populations could be attributed to reduced gene flow (Furnier and Adams 1986; Zabinski 1992), unfavorable climatic conditions (Michaud *et al.* 1995) and small effective population size (Hoffman and Blows 1994; Raijmann *et al.* 1994). One plausible explanation of the contradictory evidence could be the application of the term "marginal" in some cases to populations that were merely on the edge of the species range, but were neither small in size nor reduced in their gene flow (Dessauer *et al.* 1975; Van Rossum *et al.* 1997).

---

[1] A version of this chapter has been published. Fazekas and Yeh 2001. Genome. 44: 13-22.

The study of allozyme variation has primarily advanced the current understanding of plant population genetics. In a recent simulation study of the effects of balanced polymorphisms, background selection and local selection on genetic diversity in subdivided populations, Charlesworth *et al.* (1997) suggested that F_{ST} for unlinked loci would be much lower than that for loci linked to loci under local selection. Hence, the use of F_{ST} to measure the between-population differentiation could be ambiguous when comparisons were made between genome regions with different rates of crossing over. Even in outcrossing populations, local selection on certain loci in the genome could produce large differences in allele frequency at other loci when gene flow was restricted. Charlesworth *et al.* (1997) proposed that genetic markers reflecting neutral differences be used to measure the between-population diversity because these relate most directly to the coalescence times. They further suggested that data on several loci should be used. For "if there are local selection pressures, differences between populations are not independent of map positions, and very high levels [of differentiation] can be produced at loci in the region of a locus subject to local selection".

Lodgepole pine (*Pinus contorta* subsp. *latifolia*) is an important component of forest ecosystems in western North America (Critchfield 1980). Allozyme studies of this conifer revealed that the majority (> 90%) of the genetic diversity resides within populations (Yeh and Layton, 1979; Dancik and Yeh, 1983; Wheeler and Guries 1982). The apparently small among-populations genetic differentiation in lodgepole pine is a general feature found in most temperate and boreal conifers (Yeh 1989). Cwynar and MacDonald (1987) noted that because lodgepole pine followed macroenvironmental gradients during post-glacial migration, it is "difficult to distinguish between the relative importance of selection along these gradients and the effects resulting from the migration process itself." A stepwise model of repeated founding events may explain reductions in allelic diversity or increased differentiation in populations toward the periphery of a species, particularly those involved in recent migration events. Random events, and selective forces in marginal areas, as well as the consequences of selection during the migration event itself, may also be important, in addition to the effect of the mating system. Repeated founder effects in conjunction with directional selection for long distance dispersal may help explain a reduction in diversity of marginal populations.

Random amplified polymorphic DNA (RAPD) markers are well suited for population genetic analysis in that they are primarily neutral and highly variable. Each random primer typically yields several variable RAPDs; and since many random primers are available, a large number of RAPDs are available for analysis. RAPDs have been shown to inherit in a biparental dominant Mendelian manner (Carlson *et al.* 1991; Roy *et al.* 1992; Heun and Helentjaris 1993). Their use as markers for the genetic characterization of populations has been well established for a variety of organisms (Chalmers *et al.* 1992; Huff *et al.* 1993; Yeh *et al.* 1995; Caccone *et al.* 1997; Hogbin *et al.* 1997; Esquibet *et al.* 1998; Lou *et al.* 1998; Suazo *et al.* 1998). Using RAPDs in this study, we compare the diversity in 15 natural populations of *Pinus contorta* subsp. *latifolia*. RAPDs are particularly well suited to studying conifers because the availability of haploid material from the megagametophyte of conifer seed avoids the problems associated with interpreting dominant markers (i.e., distinguishing heterozygotes). An additional benefit in using RAPDs is that our laboratory has the map positions of over 500 RAPDs in *P. contorta* subsp. *latifolia* (Li 1998), which can be used to provide relatively unbiased estimates of RAPD diversity across the genome. Our objective was to analyse single-locus and multilocus structure in four marginal, three intermediate and eight central populations of *Pinus contorta* subsp. *latifolia*, with the purpose of determining whether marginal populations exhibit comparatively lower genetic diversity, and of identifying whether the populations sampled have significant multilocus associations.

Materials and methods

Population sampling

Seeds were collected from 15 lodgepole pine populations in British Columbia and the Yukon (Figure 2-1), with sampling taken from eight to twelve trees per population. Megagametophytes of six seeds were analysed from each tree to allow the maternal genotype to be inferred. Populations were classified into marginal, intermediate and central categories, all based on population size and geographic isolation. Populations 34, 35, 68, and 69 are small and located towards the northern (34, 35) and eastern (68, 69) limits of the species range. Populations 31, 30, and 67 are intermediate in size and are in proximity to other stands. Central populations 36, 28, 27, 60, 61, 18, 57, and 16 are large and continuous, and located in central areas of the species range. Serotinous cones were collected under closed stand conditions from trees a minimum of 150 feet (45.7m) apart, and only if cones were younger than seven years (determined by the number of nodes from the branch tip). Cone lots were kept separate to maintain the individuality of trees in the population. Populations sampled appeared to have a uniform physiography and were representative of the locale. Cones were air dried for up to six months, after which they were opened by being heated in a kiln for 12 hours at 57°C. After being removed from the cones and cleaned, the seed was stored at –20°C.

DNA extraction and amplification

Prior to DNA extraction, the seeds were re-hydrated overnight in distilled water. Seed coats were removed, and the haploid female megagametophyte was separated from the embryo. Total DNA was extracted from each megagametophyte using a phenol/chloroform method modified from that of Lee and Taylor (1990). DNA concentration was quantified on a 0.7% agarose gel by comparison with a λ-DNA standard (Roche). Working samples were diluted to 0.2 – 0.4 ng/μl. PCR amplification of DNA was carried out in a GeneAmp 9600 Thermo-cycler (Perkin-Elmer). Amplification reactions contained: 10 mM reaction buffer (Perkin-Elmer), 0.2 mM each of dATP, dCTP, dGTP, and dTTP, 2.5 mM $MgCl_2$, 0.2 μM primer, 0.6 units of Taq DNA polymerase (Perkin-Elmer), 2-4 ng sample DNA, in a total volume of 20 μl. The PCR reaction temperature profile was as follows: 92°C for 2 min.; 40 cycles of 92°C for 30s, 36°C for 30s, 72°C for 1 min., followed by 72°C for 8 min. Amplification products were loaded into a 1.4% agarose gel and electrophoresis was performed at 70V for 5 hours. Gels were stained with ethidium bromide and photographed under UV illumination.

The molecular sizes of RAPD fragments were estimated by comparison with DNA molecular weight marker VI (Roche) using DNA ProScan program (1997). Samples that had amplification failures or unscorable fragments were treated as missing data.

There has been some criticism of the use of RAPDs because they are sensitive to experimental conditions (Smith *et al.* 1994; Harris 1995). We have, however, found that exercising careful control to use a consistent quantity of DNA per amplification (2-4 ng) and using high quality *Taq* Polymerase has resulted in clear, reproducible RAPD profiles. Primer screening is also essential to obtain reproducible RAPD profiles. Two hundred ten-nucleotide random primers from the University of British Columbia Biotechnology Laboratory were initially screened on six megagametophytes from each of five populations. Ten primers that consistently revealed sharp and reproducible RAPDs over several independent runs were selected for this population study. The ten primers varied in G/C content from 50 to 70 percent (Table 2-1).

The inheritance of RAPDs in *P. contorta* subsp. *latifolia* was previously interpreted from segregation patterns of 60 to 90 megagametophytes of seeds from each of three maternal trees (Li 1998). Chi-square analysis indicated that 53 of 803 (7%) markers distorted significantly from a 1:1 Mendelian ratio. Due to their random nature, RAPDs are often assumed to generate markers from across the genome. Research on RAPD genomic maps by Li (1998) showed that markers generated from the same primers chosen in this study were indeed located on different linkage groups across the genome.

Data Analysis

The RAPDs were scored as present (1) or absent (0) for individual megagametophytes. RAPD bands are denoted here by the primer followed by a letter, starting with the smallest scorable fragment. For example, fragment 211-A refers to the smallest scorable fragment generated by primer 211. To estimate the maternal genotype, six megagametophytes per tree were analyzed for variation of RAPDs. Assuming no segregation distortion, it is expected that $(0.5)^{6-1} = 0.03$ of heterozygotes would misclassify as homozygotes for one or the other RAPD phenotype using this number of megagametophytes per tree.

POPGENE 1.31 (Yeh *et al.* 1996) was used to analyse the data. Single-locus analyses of the genotype data included estimates of heterozygosity from the equation (Nei $H_e = 1 - \sum_i^n p_i^2$ 1973), where p_i is the frequency of the ith allele, and fixation indices (Wright 1922) were calculated as: $F = 1 - (H_o/H_e)$ for each population. Differentiation between populations was estimated using F-statistics (Wright 1965). Wright's F statistics are based on the correlation between uniting gametes relative to random mating within a population, or a group of populations. Wright related the average observed heterozygosity in populations (H_I) to the average expected heterozygosity in populations (H_S) and to the average expected heterozygosity if all populations were grouped as one (H_T). He then defined three statistics to measure the departure from Hardy-Weinberg expectations.

$$F_{IS} = \frac{H_S - H_I}{H_S}$$

$$F_{IT} = \frac{H_T - H_I}{H_T}$$

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

F_{IS} measures the inbreeding in individuals relative to their population whereas F_{IT} is the relative amount of inbreeding in individuals relative to all populations pooled together. F_{ST} characterizes the amount of differentiation between populations,or the reduction in heterozygosity of individual populations relative to all populations pooled together, as a result of genetic drift.

Average gene flow (Slatkin and Barton 1989) and G^2-tests for heterogeneity of RAPD frequencies across populations were also calculated. The multilocus analyses compared the pairwise distribution of gametes using haploid data and following the methods of Brown *et al*. (1980), Brown and Feldman (1981) and Yeh *et al*. (1994). Both single-locus and multilocus analyses were carried out for each of the three population categories and for the 15 populations overall. Due to missing data (see Table 2-2), some polymorphic RAPDs were excluded from the multilocus analysis. Excluded RAPDs were 243-D, 243-E, 243-F in the marginal group, and 280-A in the central group.

The classification of populations into marginal, intermediate or central groups was based on geographic location and the size of the stand. To give further support to the classification the average genetic distance of each population to all the others was calculated. The populations were then put into their respective groups and the genetic distances averaged for comparison between groups.

The Ewens-Watterson test for neutrality (Manley 1985) was performed to determine whether the RAPDs selected for study did behave as neutral markers.

Results

RAPD Variation Within Populations

The 15 populations in this study were grouped into central, intermediate and marginal categories. Both single-locus and multilocus parameters were examined to compare the relative effect of geographic position. The ten random primers chosen for analysis generated 52 RAPDs that ranged in size from 390 to 2710 bp. Each random primer amplified between three and seven RAPDs; the grand average for the 15 populations was 5. Figure 2-2 shows a RAPD profile generated by UBC primer 635. The Ewens-Watterson test for neutrality (results not shown) indicated that for both the haploid and diploid data sets, three of the 52 RAPDs (211-D, 243-C, 250-C) were outside the lower 95% confidence limits, suggesting that they were non-neutral.

Table 2-2 gives RAPD frequencies of polymorphic RAPDs for each population and G^2-tests for heterogeneity of RAPD frequencies across populations. Thirty-two of the 52 (61.5%) RAPDs were polymorphic overall, with 19-24 polymorphic RAPDs in any one population. When considered in groups, the number of polymorphic loci increased from 26 in the marginal populations to 30 in the central group.

Three RAPDs (243-F, 250-C, and 428-E) were present at a frequency of 0.100 or less in some populations of each category. In four cases, a rare RAPD (623-F in population 68; 250-A in population 31; 428-A in population 61; and 623-B in population

18) was restricted to a single population. The frequencies at 14 of the 32 RAPDs (44%) were heterogeneous across populations suggesting that RAPD divergence among natural populations of *P. contorta* subsp. *latifolia* was due mainly to frequency differences rather than the fixation of locally common or rare RAPDs.

Although gene diversity expressed as heterozygosity, as well as Wright's fixation index are both quite variable within the three groups, there is a clear trend of decreasing diversity and increasing levels of fixation, from central to intermediate to marginal populations (Table 2-3).

RAPD Differentiation Among Populations

The F_{ST} values (Table 2-4) showed the greatest population differentiation at marginal sites. F_{IS} values increase consistently from central, to intermediate, to marginal populations. In all groups, negative values of F_{IS} suggest there is an excess of heterozygotes relative to Hardy-Weinberg expectations. That the F_{IT} value across all populations is close to zero indicates that the level of inbreeding overall is small. The overall value of $F_{ST} = 0.0981$, indicates that more than 90% of the total variation is contained within populations. Within the marginal group the average genetic distance to all other populations (including intermediate and central) was 0.041. For the intermediate and central groups this value was 0.030 and 0.028 respectively.

Multilocus analysis to identify gametic disequilibrium based on the methods of Brown and Feldman (1980) and Yeh *et al*. (1994) showed the observed variances of number of heterozygous loci (RAPDs) between two randomly chosen gametes in a population (M_2) exceeded their upper 95% confidence limits (Table 2-5). The observed third (M_3) and fourth moments (M_4) of number of heterozygous loci between two randomly chosen gametes in a population also exceeded their upper or lower 95% confidence limits in all populations, except for the third moment in Population 67 (Table 2-5). Using the method of Brown *et al.* (1980, 1981), the total and average variance in the number of heterozygous RAPDs (loci) were partitioned into the single and two-locus effects. The variance components across all populations (Table 2-6) indicate that 76% of the total variance was due to single-locus effects, while two-locus effects accounted for 24%. The most prominent component was mean gene diversity (MH), which accounted for 68% of the variation.

When split into the categories of marginal, intermediate and central, the mean gene diversity and variance of diversity are highest in the central populations and lowest in the marginal ones (Table 2-6). Of the two-locus effects, the variance of disequilibrium is highest in the central sites, whereas the two-locus Wahlund's effect is lowest in the marginal populations.

Discussion

Genetic diversity may be present in natural populations due to the effects of genetic drift on selectively neutral mutations in finite populations (Kimura 1983). Nonetheless, it is well recognized that various kinds of selection processes can also influence genetic diversity in natural populations (Gillespie 1994). One problem is that there are very few good empirical data sets available to compare with the theoretical predictions. Consequently, the level of genetic diversity in natural populations and the forces that maintain the diversity have been subjects of great debate in population genetics.

A decrease of genetic variability in marginal or disjunct populations relative to central ones has been documented (Avise and Selander 1972; Yeh and Layton 1979; Linhart and Premoli 1994; Allen *et al.* 1996), and is of concern for conservation of threatened species (Barrett and Kohn 1991; Ellestrand and Elam 1993). This decrease of variation in marginal populations can be due to the mating system, selection, and stochastic forces. Because the importance of each of these factors will depend on the species involved and on the degree of interaction between the forces, the relative magnitudes of each of the factors are difficult to differentiate.

In this study of *P. contorta* subsp. *latifolia* the most significant finding was the reduced RAPD heterozygosity (Table 2-3) and greater among-population RAPD differentiation in the marginal populations (Table 2-4). Levels of heterozygosity are highest in the central group, which also shows the lowest levels of inbreeding. Taken as a group, the marginal populations also have more than 13% fewer polymorphic RAPDs than central ones (Table 2-2). This is in agreement with an allozyme study that showed a trend of decreasing genetic variability at marginal populations in *P. contorta* subsp. *latifolia* (Yeh and Layton 1979). Values for F_{ST} indicate that populations in marginal sites are somewhat more differentiated than populations in central or intermediate areas (Table 2-4). The division of genetic diversity within and among populations, as well as estimates of heterozygosity of lodgepole pine based on RAPD markers (overall $h = 0.17$), is consistent with the estimates from other conifers (Hamrick and Godt 1990).

In all groups, there is heterozygote excess relative to Hardy-Weinberg expectations (Table 2-4) analogous to other studies in conifers that reported heterozygote excess based on allozymes (Yeh *et al.* 1986; Perry *et al.* 1990; Rajora *et al.* 1998). However, this result contrasts with a previous allozyme study of *P. contorta* subsp. *latifolia* sampled in Alberta that showed a slight deficiency of heterozygotes (Dancik and Yeh 1983). A number of variables could contribute in part to heterozygote excess in plants. Heterozygote advantages of the marker loci, or loci tightly linked to the markers, could result in a selective force leading to an excess of heterozygote. This cause may only be partly responsible, given that the results of the Ewens-Watterson test, which showed that 49 of 52 RAPDs were neutral in this study. Nevertheless, the three RAPDs (211-D, 243-C, 250-C) that were non-neutral were outside the lower 95% confidence limits in the Ewens-Watterson test. This signifies that the frequencies of the markers at these RAPD loci were too even, suggesting that there was a tendency in favour of heterozygotes in the population. Negative assortative mating is another possible explanation for heterozygote excess. If flowering phenology or timing promoted pollination by genotypically different trees occurs, a frequency of heterozygotes in excess of H-W expectations could result. Isabel *et al.* (1995) however, argued that the slight observed excess of heterozygotes in *Picea mariana* (Mill.) often seen in mature populations of conifers might be a result of selection against inbred individuals. This would affect the whole genome (including neutral loci), whereas balancing selection would affect only particular loci conferring heterozygote advantage.

The estimate of M_2 exceeded the upper 95% confidence limits in all 15 populations (Table 2-5), indicating the significance of two-locus gametic disequilibria.

The estimates of M_3 and M_4 also exceeded their upper or lower 95% confidence limits in all populations except for the third moment in Population 67 (Table 2-5). This signified predominance of three- and four-locus gametic disequilibria in *P. contorta* subsp. *latifolia*, respectively.

The method of Brown *et al.* (1981) partitions the total and average variance in the number of heterozygous loci into the single and two-locus effects (Table 2-6). The single-locus components are dominated by mean gene diversity (MH). This component accounts for 90% of the single-locus effect and 68% of the total variance. When divided into the three groups, the single-locus components confirm the pattern of increasing diversity towards the center of the species range. Within each group, the partitioning of the various single-locus components among MH, VH (the variation among populations in gene diversity), and WH (Wahlund effect) is very similar.

The two-locus components reveal additional population structure. Over all 15 populations, the most important contributors to the two-locus effects are the variance of disequilibrium (VD), which accounts for 46%, and the two-locus Wahlund's effect (WC), which explains 32% of the variation. A high value of VD means that the most common gametic types are dissimilar in different populations. This is an indication that the multilocus associations present are largely explained by founder effects (Brown and Feldman 1981). As regards the three categories, it can be seen that VD, as a percentage of the two-locus effects, is a strong component in each group, although the values decrease towards the central populations. Founder effects are generally attributed to small effective population size. This attribution is consistent with our data, which show that the greatest founder effects appear in the marginal sites. *P. latifolia* subsp. *latifolia* forests typically arise from fire events, with seeds being released from serotinous cones (Critchfield 1980). This natural system of regeneration permits severe genetic bottlenecks, which should result in the presence of large amounts of gametic disequilibria (Yang and Yeh 1993). Repeated fire events could result in a continuous series of founder effects. This may explain the high values of VD in the central and intermediate populations. These populations may not have enough time for gametic disequilibria to decay before another fire event occurs.

These data also fit the model of stepwise founding events presented by Cwynar and MacDonald (1987). As lodgepole pine migrated northward following deglaciation, repeated founder events followed by dispersal would result in populations on the leading edge of the migration to have lower levels of diversity as well as significant gametic disequilibria. Over time this disequilibria would decay. However, as established populations grew and interconnected, they would be increasingly at risk from fire events that would allow gametic disequilibria to persist.

The two-locus Wahlund's effect (WC) is characteristic of strong population subdivision (Brown and Feldman 1981). It measures the pooling of genes from different populations that have different multilocus associations. Although long distance pollen dispersal allows central, physically closer populations to have greater gene flow (Table 2-4) and to exchange genes more frequently than marginal ones, WC is greatest in the central populations. The high level of gene flow has not resulted in multilocus equilibrium. Repeated founder effects may have allowed different multilocus associations to be maintained between populations. Table 2-2 also reflects this possibility: there is a large amount of heterogeneity among allele frequencies, and some alleles (e.g., at loci 280-525, 623-1220) are present in only a few populations, but at significant frequencies. The isolation and comparatively smaller size of the marginal populations can cause increased inbreeding and genetic drift to occur, thereby increasing the degree of differentiation between populations. Despite strong differences between the marginal populations (Table 2-4), the distances between them are too great to permit a large amount of gene flow. Additionally, if there are fewer polymorphic loci, the opportunities for linkage disequilibrium to be detectable will be less. The total amount of variation in both the single-locus and multilocus components increase from marginal to central populations.

AI is the interaction effect of WC x MD. A positive value of AI indicates that the correlation of alleles between populations is similar to the pattern within populations (Brown and Feldman 1981). AI is considerably larger in the marginal populations. This may be due in part to the marginal populations' having fewer polymorphic loci, from which potential two-locus associations may occur. The values for CI, which represents the covariation in the interaction of disequilibria and Wahlund's covariance among populations (Brown and Feldman 1981), are lowest in the central group. High values of CI are also associated with founder effects, which further supports the values given by the variance of disequilibrium (VD).

Multilocus structure is important when there is gametic disequilibrium or nonrandom association of nonalleles between physically independent as well as linked loci (Barton and Clark 1990). Empirical studies showed that gametic disequilibrium was the result of many evolutionary mechanisms, including epistatic selection (Brown *et al.* 1977), chance from population subdivision and founder effect (Waller and Knight 1989) and genetic hitchhiking (Laurie-Ahlberg ~~and~~ Weir 1979). The prediction for outcrossing plants such as conifers has been that any initial gametic disequilibria will decay rapidly without strong epistatic selection. Nevertheless, this is a prediction of long-term behavior. In the establishment of forests after a fire, for example, there could initially be gametic disequilibria that slowly decays for closely-linked genes, or if there were nonrandom mating. This consideration is possible for *P. contorta* subsp. *latifolia* since most of its forests are thought to be of fire origin (Critchfield 1980). Significant multilocus structures have also been reported in other tree species (Boyle 1985; Muona and Szmidt 1985; Roberds and Brotchol 1985; Yeh and Morgan 1987; Yeh *et al.* 1994; Kremer and Zanetto 1997), including *P. contorta* subsp. *latifolia* (Yang and Yeh 1993). Epistatic selection, founder effect and population subdivision were cited as the major causes for the observed gametic disequilibria.

Investigations into the multilocus structure of populations, in conjunction with single-locus analyses are important for revealing information otherwise unavailable. The single-locus analysis here clearly reinforces the concept of reduced genetic diversity in marginal populations of *P. contorta* subsp. *latifolia* relative to central ones of the same species. Additionally, the multilocus analysis revealed that founder effects and multilocus Wahlund's effect are two prominent forces contributing to the genetic structure of populations of this conifer. The primary concern in conservation of populations or a species is population size and levels of gene flow between populations. Small or isolated populations located at the limits of the species range are at greater risk from genetic drift and higher levels of inbreeding, which can result in fixation and loss of rare alleles. Central populations of *P. contorta* subsp. *latifolia* are at the least amount of risk from such forces. With large population size and extensive gene flow, their genetic diversity can be maintained; however, sampling strategies for gene conservation in these areas may need revision. Despite low F_{ST}, the large two-locus Wahlund's effect indicates that allelic combinations exist in different frequencies in different populations. Thus, sampling strategies will need to take this into account when it is desirable to capture the various multilocus associations that exist. The reduced genetic resources of marginal populations may also need some attention. The neutral markers employed in this study indicate that the reduction in heterozygosity is not the result of selective forces but is likely due to reduced levels of gene flow and to small population size. Care will need to be taken to ensure that adequate (natural) levels of gene flow are maintained in marginal populations of *P. contorta* subsp. *latifolia*, which are at risk from human disturbance. It is prudent to see whether investigations based on neutral markers in marginal populations of other species may also reveal similar results, especially in highly outcrossing species.

References

Allen, G.A., J.A. Antos, A.C. Worley, T.A Suttil and R.J. Hebda. 1996. Morphological and genetic variation in disjunct populations of the avalanche lily *Erythronium montanum*. Can. J. Bot. **74**: 403-412.

Arroyo, M.T.K. 1975. Electrophoretic studies of genetic variation in natural populations of allogamous *Limnanthes alba* and autogamous *Limnanthes floccosa* (Limnanthaceae). Heredity **35**: 153-164.

Avise, J.C. and R.K. Selander. 1972. Evolutionary genetics of cave-dwelling fishes of the genus *Astyanax*. Evolution **26**: 1-9.

Barrett, S.C.H. and J.R. Kohn. 1991. Genetic and evolutionary consequences of small population size. *In*: Genetics and Conservation of Rare Plants. Edited by D.A. Falk and K.E. Holsinger. Oxford University Press, New York. pp. 3-30.

Barton, N. and A. Clark. Population structure and processes in evolution. *In*: Population Biology. Edited by K. Wöhrmann and S.K. Jain. Springer-Verlag, Berlin. pp. 115-173.

Boyle, T.J.B. 1985. The mating system and population structure of black spruce in central New Brunswick and its influence on tree improvement strategy. PhD. Thesis. University of New Brunswick.

Brown, A.H.D., E. Nevo and D. Zohary. 1977. Association of alleles at esterase loci in wide barley *Hordeum spontaneum*. Nature **268**: 430-431.

Brown, A.H.D., M.W. Feldman and E. Nevo. 1980. Multilocus structure of natural populations of *Hordeum spontaneum*. Genetics **96**: 523-536.

Brown, A.H.D. and M.W. Feldman. 1981. Population structure of multilocus associations. Proc. Natl. Acad. Sci. **78**: 5913-5916.

Caccone, A., G. Allegrucci, C. Fortunato and V. Sbordoni. 1997. Genetic differentiation within the European sea bass (*D. labrax*) as revealed by RAPD-PCR assays. J. Hered. **88**: 316-324.

Carlson, J.E., L.K. Tulsieram, J.C. Glaubitz, V.W.K. Luk, C. Kauffeldt and R. Rutledge. 1991. Segregation of random amplified DNA markers in F_1 progeny of conifers. Theor. Appl. Genet. **83**: 194-200.

Chalmers, K.J., R. Waugh, J.I. Sprent, A.J. Simons and W. Powell. 1992. Detection of genetic variation between and within populations of *Gliricidia sepium* and *G. maculata* using RAPD markers. Heredity **69**: 465-472.

Charlesworth, B., M. Nordborg and D. Charlesworth. 1997. The effect of local selection, balanced polymorphism, and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res., Camb. **97**: 175-181.

Critchfield, W.B. 1980. Genetics of Lodgepole Pine. U.S. For. Serv. Res. Pap. WO-37.

Cwynar, L.C. and G.M. MacDonald. 1987. Geographical variation of lodgepole pine in relation to population history. Am. Nat. **129**: 463-469.

DNA Proscan Inc. 1997. Pro-RFLP Version 2.34. Nashville, Tennessee.

Dancik, B.P. and F.C. Yeh. 1983. Allozyme variability and evolution of lodgepole pine (*Pinus contorta* var. *latifolia*) and jack pine (*P. banksiana*) in Alberta. Can. J. Genet. Cytol. **25**: 57-64.

Dessauer, H.C., E. Nevo and K.-C. Chuang. 1975. High genetic variability in an ecologically variable vertebrate, *Bufo viridis*. Biochem. Genet. **13**: 651-661.

Ellestrand, N.C. and D.R. Elam. 1993. Population genetic consequences of small population size: implications for plant conservation. Annu. Rev. Ecol. Syst. **24**: 217-242.

Esquibet, M., S. Bekal, P. Castagnone-Sereno, J.P. Gauthier, R. Rivoal and G. Caubel. 1998. Differentiation of normal and giant *Vicia faba* populations of the stem nematode *Ditylenchus dipsaci*: agreement between RAPD and phenotypic characteristics. Heredity **81**: 291-298.

Furnier, G.R. and W.T. Adams. 1986. Geographic patterns of allozyme variation in Jeffrey pine. Am. J. Bot. **73**: 1009-1015.

Gillespie, J.H. 1994. Alternatives to the neutral theory. *In*: Non-neutral evolution – Theories and molecular data. Edited by B. Golding. Chapman and Hall, Toronto. pp. 1-17.

Hamrick, J.L. and M.J. Godt. 1990. Allozyme diversity in plant species. *In*: Plant population genetics, breeding and genetic resources. Edited by A.H.D. Brown, M.T. Clegg, A.L. Kahler, and B.S. Weir. Sinauer Associates, Inc., Sunderland, Mass. pp. 43-63.

Harris, S.A. 1995. Systematics and randomly amplified polymorphic DNA in the genus *Leucaena* (Leguminosae, Mimosoideae). Pl. Syst. Evol. **197**: 195-208.

Heun, M. and T. Helentjaris. 1993. Inheritance of RAPDs in F_1 hybrids of corn. Theor. Appl. Genet. **85**: 961-968

Hoffmann, A.A. and M.W. Blows. 1994. Species borders: ecological and evolutionary perspectives. Trends Ecol. Evol. **9**: 223-227.

Hogbin, P.M., D.J. Ayre and R.J. Whelan. 1998. Genetic variation and reproductive success of road verge populations of the rare shrub *Grevilla barklyana* (Proteaceae). Heredity **80**: 180-186.

Huff, D.R., R. Peakall and P.E. Smouse. 1993. RAPD variation within and among natural populations of outcrossing buffalograss [*Buchloe dactyloides* (Nutt.) Engelm.]. Theor. Appl. Genet. **86**: 927-934.

Isabel, N., J. Beaulieu and J. Bousquet. 1995. Complete congruence between gene diversity estimates derived from genotypic data at enzyme and random amplified polymorphic DNA loci in black spruce. Proc. Natl. Acad. Sci. **92**: 6369-6373.

Kremer, A. and A. Zanetto. 1997. Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. II: Multilocus patterns of diversity. Heredity **78**: 476-489.

Kimura, M. 1986. DNA and the neutral theory. Phil. Trans. Roy. Soc. Lond. **312**: 343-354.

Laurie-Ahlberg, C.C. and B.S. Weir. 1979. Allozymic variation and linkage disequilibrium in some laboratory populations of *Drosophila melanogaster*. Genetics **92**: 1295-1314.

Lee, S.B. and J.W. Taylor. 1990. Isolation of DNA from Fungal Mycelia and Single Spores. *In*: PCR Protocols. Edited by M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White. Academic Press Inc., San Diego, Cal. pp. 282-287.

Levin, D.A. 1977. The organization of genetic variability in *Phlox drummondii*. Evolution **31**: 477-494.

Li, Changxi. 1998. Genetic dissection of resistance to western gall rust (*Endocronartium harknessii* (J. P. Moore) Y. Hiratsuka) in lodgepole pine (*Pinus contorta* var.

latifolia Engelm.). Ph.D. Thesis. University of Alberta. Edmonton, Alberta, Canada.

Linhart, Y.B. and A.C. Premoli. 1994. Genetic variation in central and disjunct populations of *Lilium parryi*. Can. J. Bot. **72**: 79-85.

Lou, K.F., M.J. Weiss, P.L. Bruckner, W.L. Morrill, L.E. Talbert and J.M. Martin. 1998. RAPD variation within and among geographic populations of wheat stem sawfly (*Cephus cinctus* Norton). J. Hered. **89**: 329-335.

Loveless, M.D. and J.L. Hamrick. 1984. Ecological determinants of genetic structure in plant populations. Ann. Rev. Ecol. Syst. **15**: 65-95.

Manly, B.F.J. 1985. The Statistics of Natural Selection. Chapman and Hall, London.

Mayr, E. 1970. Populations, Species and Evolution. Harvard University Press, Cambridge, MA.

Michaud, H., L. Toumi, R. Lumaret, T.X. Li, F. Romane and F. Di Guisto. Effect of geographical discontinuity on genetic variation in *Quercus ilex* L. (holm oak). Evidence from enzyme polymorphism. Heredity **74**: 590-606.

Muona, O. and A.E. Szmidt. 1985. A multilocus study of natural populations of *Pinus sylvestris*. *In*: Population Genetics in Forestry. Edited by H.R. Gregorius. Springer-Verlag, Berlin. pp. 241-252.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Am. Nat. **106**: 283-292.

Perry, D.J., P. Knowles and F.C. Yeh. 1990. Allozyme variation of *Thuja occidentalis* L. in Northwestern Ontario. Biochem. Syst. Ecol. **18**: 111-115.

Prakash, S. 1973. Patterns of gene variation in central and marginal populations of *Drosophila robusta*. Genetics **75**: 347-369.

Raijmann, L.E.L., N.C. Van Leeuwen, R. Kersten, J.G.B. Oostermeijer, H.C.M. Den Nijs and S.B.J. Menken. 1994. Genetic variation and outcrossing rate in relation to population size in *Gentiana pneumonanthe* L. Conserv. Bio. **8**: 1014-1026.

Rajora, O.P., L. De Verno, A. Mosseler and D.J. Innes. 1998. Genetic diversity and population structure of disjunct Newfoundland and central Ontario populations of eastern white pine (*Pinus strobus*). Can. J. Bot. **76**: 500-508.

Roberds, J.H. and J.V. Brotschol. 1985. Linkage disequilibrium among allozyme loci in natural populations of *Liriodendron tulipifera* L. Silvae Genet. **34**: 137-141.

Roy, A., N. Frascaria, J. Mackay and J. Bousquet. 1992. Segregating random amplified polymorphic DNAs (RAPDs) in *Betula alleghaniensis*. Theor. Appl. Genet. **85**: 173-180.

Slatkin, M. and N.H. Barton. 1989. A comparison of three indirect methods for estimating average levels of gene flow. Evolution **43**: 1349-1368.

Smith, J.J., J.S. Scott-Craig, J.R. Leadbetter, G.L. Bush, D.L. Roberts and D.W. Fulbright. 1994. Characterization of random amplified polymorphic DNA (RAPD) products from *Xanthomonas campestris* and some comments on the use of RAPD products in phylogenetic analysis. Mol. Phyl. Evol. **3**: 135-145.

Suazo, A., R. McTiernan and H.G. Hall. 1998. Differences between African and European honey bees (*Apis mellifera* L.) in random amplified polymorphic DNA (RAPD). J. Hered. **89**: 32-36.

Van Rossum, F., X. Vekemans, P. Meerts, E. Gratia and C. Lefèbvre. 1997. Allozyme variation in relation to ecotypic differentiation and population size in marginal populations of *Silene nutans*. Heredity **78**: 552-560.

Waller, D.M. and S.E. Knight. 1989. Genetic consequences of outcrossing in the cleistogamous annual, *Impatiens capensis*: III. Interlocus associations. Heredity **63**: 1-9.

Wheeler, N.C., and R.P. Guries. 1982. Biogeography of lodgepole pine. Can J. Bot. **60**: 1805-1814.

Wright, S. 1922. Coefficients of inbreeding and relationship. Am. Nat. **56**: 330-338.

Wright, S. 1965. The interpretation of population structure by *F*-statistics with special regards to systems of mating. Evolution **19**: 395-420.

Xie, C.Y., B.P. Dancik and F.C. Yeh. 1992. Genetic structure of *Thuja orientalis*. Biochem. Syst. Ecol. **20**: 433-441.

Yang, R.-C. and F.C. Yeh. 1993. Multilocus structure in *Pinus contorta* Dougl. Theor. Appl. Genet. **87**: 568-576.

Yeh, F.C. and C. Layton. 1979. The organization of genetic variability in central and marginal populations of lodgepole pine *Pinus contorta* spp. *latifolia*. Can. J. Gen. Cyt. **21**: 487-503.

Yeh, F.C., M.A.K. Khalil, Y.A. El-Kassaby and D.C. Trust. 1986. Allozyme variation in *Picea mariana* from Newfoundland: genetic diversity, population structure, and analysis of differentiation. Can. J. For. Res. **16**: 713-720.

Yeh, F.C. and K. Morgan. 1987. Mating system and multilocus associations in a natural population of *Pseudotsuga menziesii* (Mirb.) Franco. Theor. Appl. Genet. **73**: 799-808.

Yeh, F.C. 1989. Isozyme analysis for revealing population structure for use in breeding strategies. *In*: Breeding tropical trees: Population structure and genetic improvement strategies in clonal and seedling forestry. Proc. IUFRO Conference, Pattaya, Thailand, Nov. 1988. Edited by G.L. Gibson, A.R. Griffin, and A.C. Matheson. Oxford Forestry Institute, Oxford and Winrock International, Arlington, Virg. pp. 119-131.

Yeh, F.C., J. Shi, R.-C. Yang, J. Hong and Z. Ye. 1994. Genetic diversity and multilocus associations in *Cunninghamia lanceolata* (Lamb.) Hook from The People's Republic of China. Theor. Appl. Genet. **88**: 465-471.

Yeh, F.C., D.K.X. Chong and R.-C. Yang. 1995. RAPD variation within and among natural populations of trembling aspen (*Populus tremuloides* Michx.) from Alberta. J. Hered. **86**: 454-460.

Yeh, F.C., R.-C. Yang, J. Mao, Z. Ye and T.J.B. Boyle. 1996. POPGENE, the Microsoft Windows-based user-friendly software for population genetic analysis of co-dominant and dominant markers and quantitative traits. Department of Renewable Resources, University of Alberta, Canada.

Zabinski, C. 1992. Isozyme variation in eastern hemlock. Can. J. For. Res. **22**: 1838-1842.

Table 2-1: Sequences of the University of British Columbia (UBC) Biotechnology Laboratory decamer primers used in this study of *Pinus contorta* subsp. *latifolia*.

UBC Primer	Sequence
211	5' -GAAGCGCGAT- 3'
232	5' -CGGTGACATC- 3'
243	5' -GGGTGAACCG- 3'
250	5' -CGACAGTCCC -3'
251	5' -CTTGACGGGG- 3'
280	5' -CTGGGAGTGG- 3'
428	5' -GGCTGCGGTA- 3'
429	5' -AAACCTGGAC- 3'
623	5' -TGCGGGACTG- 3'
635	5' -CTCAGCTCAG- 3'

Table 2-2: Frequencies of 33 Polymorphic RAPDs in 15 populations of *Pinus contorta* subsp. *latifolia*.

	Marginal				Intermediate			Central								
RAPD Marker Population:	34	35	68	69	31	30	67	36	28	27	60	61	18	57	16	G^2
211-B	0.682	0.727	0.625	0.583	0.625	0.778	0.583	0.722	0.688	0.708	0.556	0.708	0.800	0.563	0.708	7.5
211-C	0.909	1.000	1.000	1.000	0.938	1.000	1.000	0.778	1.000	1.000	1.000	1.000	1.000	0.875	1.000	29.4^{**}
211-D	0.546	0.182	0.417	0.583	0.688	0.500	0.500	0.667	0.313	0.375	0.444	0.458	0.400	0.563	0.583	20.8
211-E	0.773	0.864	0.708	0.542	0.500	0.833	0.667	0.667	0.688	0.500	0.667	0.583	0.800	0.625	0.667	17.3
243-A	0.636	0.636	0.792	0.417	0.625	0.611	0.583	0.389	0.750	0.667	0.556	0.625	0.600	0.625	0.625	12.4
243-B	0.773	0.773	0.583	0.667	0.688	0.778	0.958	0.778	0.750	0.667	0.722	0.792	0.550	0.563	0.625	19.4
243-C	0.455	0.591	0.625	0.458	0.438	0.611	0.500	0.611	0.750	0.667	0.500	0.583	0.500	0.375	0.500	10.6
243-D	0.773	0.864	—	1.000	0.938	1.000	0.792	1.000	1.000	1.000	1.000	0.958	0.750	1.000	0.875	36.6^{**}
243-E	0.727	0.682	—	0.708	0.375	0.778	0.875	0.556	0.500	0.417	0.667	0.750	0.500	0.563	0.667	24.1
243-F	0.046	0.273	—	0.083	0.000	0.222	0.083	0.333	0.750	0.292	0.222	0.125	0.350	0.375	0.083	39.9^{**}
250-A	1.000	1.000	1.000	1.000	0.938	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	5.9
250-C	0.546	0.682	0.042	0.125	0.438	0.390	0.292	0.389	0.688	0.750	0.833	1.000	0.650	0.500	0.333	97.9^{**}
250-D	0.773	0.318	0.708	0.708	0.563	0.611	0.708	0.611	0.438	0.583	0.111	0.708	0.400	0.563	0.667	38.8^{**}
251-B	0.455	0.727	0.833	0.667	0.688	0.667	0.458	0.611	0.643	0.667	0.500	0.458	0.300	0.625	0.583	23.4
251-C	0.818	1.000	1.000	0.875	0.813	0.611	0.833	0.944	0.929	0.917	0.722	0.750	1.000	1.000	0.958	40.8^{**}
280-A	1.000	1.000	1.000	1.000	0.813	0.938	0.708	0.889	0.688	1.000	0.833	—	1.000	1.000	1.000	49.2^{**}
428-A	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.917	1.000	1.000	1.000	10.4
428-B	0.909	0.955	0.917	1.000	0.938	1.000	1.000	1.000	0.750	0.958	1.000	0.917	1.000	1.000	1.000	25.0
428-C	0.727	0.409	0.708	0.458	0.688	0.667	0.500	0.611	0.625	0.667	0.667	0.708	0.800	0.563	0.625	14.9
428-D	0.773	0.773	0.625	0.667	0.813	0.889	0.333	0.889	0.563	0.667	0.611	0.542	0.750	0.438	0.458	35.5^{*}
428-E	0.091	0.227	0.042	0.167	0.188	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.150	0.000	0.125	33.5^{*}
429-A	1.000	0.955	0.792	0.917	1.000	1.000	0.958	1.000	0.875	0.958	0.944	0.818	1.000	0.938	0.833	25.7
429-B	0.818	0.636	0.708	0.708	0.813	0.667	0.667	0.667	0.563	0.792	0.667	0.636	0.800	0.625	0.750	8.0
429-C	1.000	1.000	1.000	1.000	1.000	0.667	1.000	1.000	0.938	1.000	1.000	1.000	0.900	1.000	1.000	38.0^{**}
623-B	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.950	1.000	1.000	5.5
623-C	1.000	1.000	1.000	0.958	0.938	1.000	0.625	0.778	0.563	0.833	0.778	0.958	1.000	0.750	1.000	62.7^{**}
623-D	0.773	1.000	0.917	0.917	0.938	1.000	1.000	0.889	0.875	0.833	0.944	0.792	0.857	1.000	1.000	27.8^{*}
623-E	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.167	0.000	0.000	0.250	0.000	0.000	33.0^{*}
623-F	1.000	1.000	0.875	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	14.3
635-A	0.917	0.864	0.792	0.792	0.688	0.944	0.917	0.833	0.750	0.875	0.556	0.600	0.800	0.688	0.750	18.8
635-B	0.500	0.636	0.708	0.708	0.625	0.611	0.458	0.611	0.875	0.792	0.833	0.900	0.750	0.625	0.708	16.8
635-C	1.000	0.818	0.833	1.000	1.000	0.778	1.000	0.778	1.000	1.000	0.813	1.000	1.000	1.000	0.833	36.6^{*}
Sample size (No. gametophytes)	66	66	72	72	48	54	72	54	48	72	54	72	60	48	72	
Polymorphic loci	22	21	20	21	24	20	21	22	23	23	22	22	22	19	21	
Polymorphic loci per group	26				28			30								

Note: G^2 test for heterogeneity of RAPD frequencies across populations: *significant at $P \leq .01$ and **significant at $P \leq .001$ respectively. Missing data denoted by —.

Table 2-3: Comparison of genetic variation and fixation indices in marginal, intermediate and central populations of *Pinus contorta* subsp. *latifolia*.

Population	megagametophytes sampled	Heterozygosity ± standard error[†]	Fixation Index	Latitude	Longitude	Elevation (m)
Marginal						
34	72	0.1464 ± 0.0265	-0.1168	60.41	136.11	747
35	66	0.1418 ± 0.0263	-0.0726	59.48	133.47	789
68	66	0.1390 ± 0.0270	-0.0741	51.01	115.02	1501
69	72	0.1446 ± 0.0274	-0.0746	49.26	114.25	1379
Mean		0.1430 ± 0.0268[a]	-0.0845			
Intermediate						
31	48	0.1582 ± 0.0276	-0.0790	61.10	129.20	884
30	54	0.1453 ± 0.0277	-0.1198	59.59	128.33	640
67	72	0.1500 ± 0.0286	-0.1113	53.16	117.09	1204
Mean		0.1512 ± 0.0280[b]	-0.1034			
Central						
36	54	0.1607 ± 0.0282	-0.0902	57.29	130.13	815
28	48	0.1706 ± 0.0287	-0.1290	58.40	124.10	762
27	72	0.1548 ± 0.0271	-0.1079	57.00	122.24	1113
60	54	0.1579 ± 0.0281	-0.2229	55.33	122.33	732
61	72	0.1698 ± 0.0300	-0.1985	53.52	121.48	838
18	60	0.1565 ± 0.0274	-0.1284	51.59	123.45	1059
57	48	0.1609 ± 0.0309	-0.1206	49.54	118.12	579
16	72	0.1530 ± 0.0282	-0.1837	51.06	121.40	1814
Mean		0.1605 ± 0.0286[c]	-0.1476			

Note: [†] Nei's (1973) expected heterozygosity. *a*, significantly different from *b* and *c* at $p<0.038$, and $p<0.0003$ respectively. *b*, significantly different from *c* at $p<0.032$.

Table 2-4: Comparison of *F*-statistics and number of migrants (*Nm*) across marginal, intermediate, and central populations in *Pinus contorta* subsp. *latifolia*.

Group	Populations per Group	F_{IS}	F_{IT}	F_{ST}	*Nm*
Marginal	4	-0.0850 (± 0.017)	0.0210 (± 0.018)	0.0977 (± 0.011)	2.30
Intermediate	3	-0.1026 (± 0.011)	-0.0204 (± 0.010)	0.0745 (± 0.010)	3.10
Central	8	-0.1484 (± 0.013)	-0.0609 (± 0.010)	0.0762 (± 0.009)	3.03
Across populations	15	-0.1114 (± 0.017)	-0.0024 (± 0.016)	0.0981 (± 0.010)	2.23

Table 2-5: The observed variance (M2), third moment (M3) and fourth moment (M4) of number of heterozygous loci between two randomly chosen gametes *Pinus contorta* subsp. *latifolia*. Values exceeding 95% confidence limits are denoted by *.

Population	M_2	M_3	M_4
34	7.7824*	-7.3451*	193.5774*
35	93.5788*	2940.8305*	111980.8595*
68	35.8991*	830.0216*	27131.4433*
69	5.9150*	-3.1431*	115.2298*
31	188.8439*	4913.4162*	188096.1488*
30	14.8197*	57.4551*	1246.3681*
67	11.7887*	3.1696	463.2211*
36	80.5964*	2158.9345*	82722.7457*
28	52.5879*	-19.8146*	7240.3206*
27	9.9183*	-16.3394*	346.5404*
60	23.4501*	45.4990*	1254.9552*
61	8.1768*	-6.3583*	197.5663*
18	8.1404*	-8.8594*	219.5746*
57	9.8189*	-6.0705*	330.1855*
16	6.1554*	-4.0396*	124.5061*

Table 2-6: Components of variance of multilocus associations in *Pinus contorta* subsp. *latifolia*.

Component	Marginal	Intermediate	Central	Across populations
Single locus effect				
Mean gene diversity (MH)	3.6490	4.1670	4.2438	3.6917
Variance of diversity (VH)	0.1580	0.2307	0.2480	0.2389
Wahlund's effect (WH)	0.1741	0.2085	0.1845	0.1888
Total	3.9811	4.6062	4.6763	4.1194
Two-locus effect				
Mean disequilibrium(MD)	0.4374	0.8114	0.7353	0.3912
Wahlund's effect (WC)	0.0661	0.1643	1.4367	0.8134
Interaction between MD and WC (AI)	0.1190	0.0207	0.0758	0.0863
Variance of disequilibrium (VD)	0.7811	1.0514	1.6809	1.1785
Covariance of interaction (CI)	0.0615	0.2449	0.0148	0.0827
Total variance (MH + VH + WH + MD + WC + AI)	4.6036	5.6025	6.9242	5.4103
Average variance (MH + MD + AI + VD + CI)	5.0481	6.2954	6.7506	5.4305

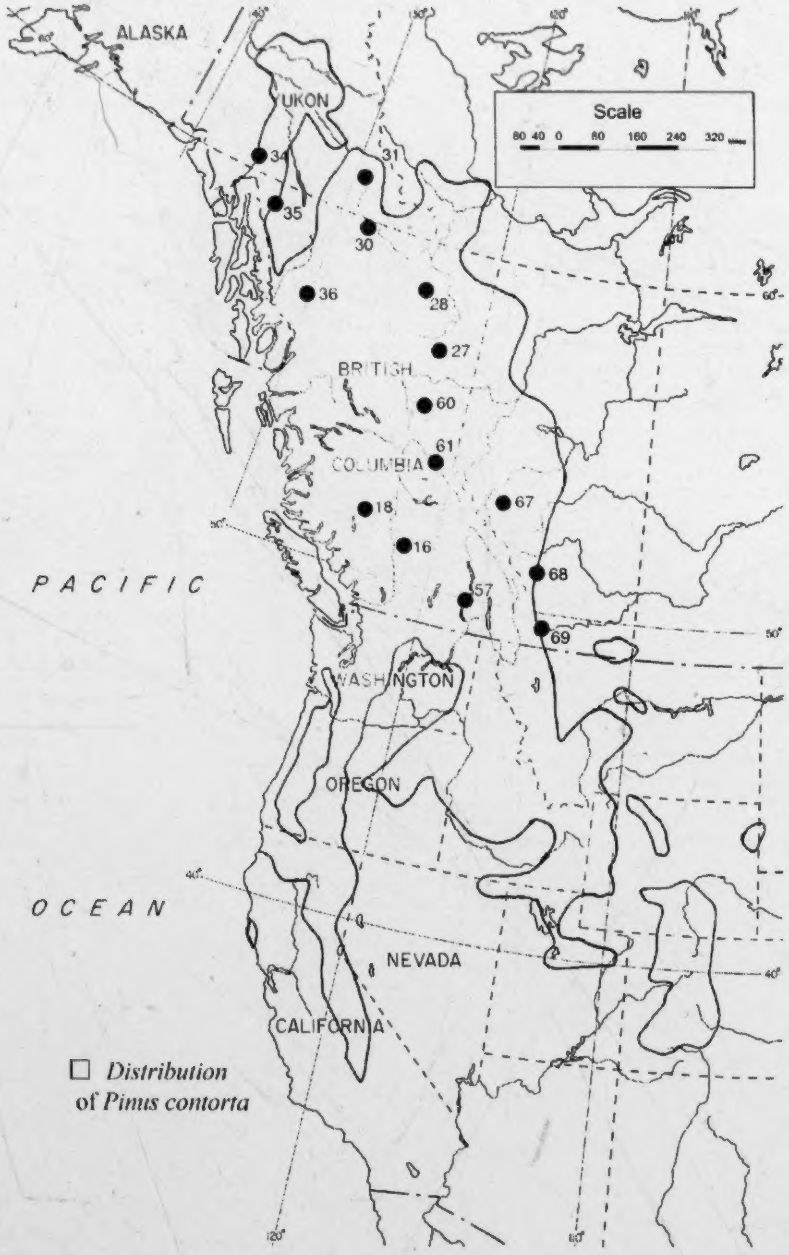
ALASKA
YUKON
Scale
80 40 0 80 160 240 320
34
35
31
30
36
28
27
BRITISH
60
61
COLUMBIA
18
67
16
68
57
69
PACIFIC
WASHINGTON
OREGON
OCEAN
NEVADA
CALIFORNIA
☐ Distribution
of Pinus contorta
140°
130°
120°
110°
60°
50°
40°

Figure 2-1: Locations of 15 Populations of *Pinus contorta* subsp. *latifolia*.

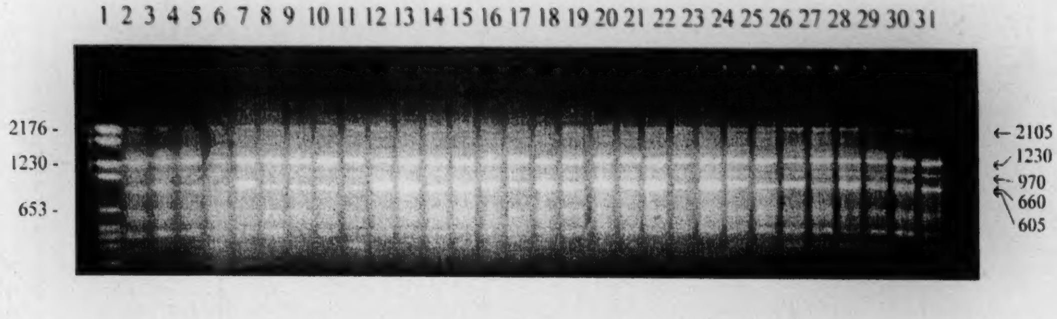
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
2176 -
1230 -
653 -
← 2105
1230
970
660
605

Figure 2-2: RAPD profiles amlified with UBC primer 635. Lane 1 is the DNA molecular weight marker VI (Boehringer Mannheim) with size in base pairs indicated on the left. Lanes 2–31 are amplified fragments of megagametophyte template DNAs. The arrows on the right represent scored fragments, with their molecular sizes in base pairs.

Chapter 3

Inferences about the post glacial history of *Pinus contorta* from gene flow

Introduction

Lodgepole pine (*Pinus contorta* Dougl.) is a major component of forest ecosystems in western North America (Critchfield 1980). It has a wide distribution, with populations extending from the Yukon to California and from the Pacific coast to east of the Rocky Mountains (Figure 3-1). Much of the present day range of the species was covered in ice during the last glaciation (the Wisconsin), which retreated over the period of approximately 14,000 –10,000 years before present (Dyke and Prest 1987). *Pinus contorta* persisted south of the ice sheet, as evidenced from the fossil record of the late glacial - early postglacial period (Baker 1970; Mack *et al.* 1978; Carrara and Wilcox 1984).

Because lodgepole pine now occupies an area in the Yukon that was unglaciated during the Wisconsin, it was suggested (Hultén 1937) that this area might have served as a northern refugium. Northern populations have been shown to be somewhat distinct. They have a high frequency of three-needled fascicles (von Rudloff and Nyland 1979; Critchfield 1980) and show differences in resin monoterpenes (von Rudloff and Nyland 1979; Forrest 1980, 1981) and allozymes (Yeh and Layton 1979; Wheeler and Guries 1982) from southern populations. These morphological and genetic differences have been used to support the hypothesis that lodgepole pine survived the last glaciation in the

Yukon. Populations that have been isolated from one another for a long period of time can diverge due to the effects of random drift and mutation.

It has also been hypothesized (Wheeler and Guries 1982) that the coastal variety of lodgepole pine (*Pinus contorta* subsp. *contorta*) survived the last glaciation in ice-free areas that may have existed in parts of the Queen Charlotte Islands (Calder and Taylor 1968), Vancouver Island, and Alexander Island (Heusser 1960).

The theory of isolation by distance (Wright 1943) predicts that genetic differentiation at neutral loci will increase with increased geographic distance. The further apart populations are geographically, the less opportunity there is for gene flow to occur between them. Slatkin (1993) suggested a method of testing for isolation by distance, and used simulations to show that this effect in fact can be detected. He defined an index of genetic similarity ($\hat{M}$) that is derived from the relationship $F_{ST} = 1/(4M+1)$ where F_{ST} is the genetic divergence between two sampled populations. For estimating gene flow between population pairs, Slatkin and Barton (1989) used G_{ST} (Nei 1973) and $\hat{\theta}$ (Weir and Cockerham 1984) as estimators of F_{ST}. Slatkin (1993) suggested that a log-log graph of $\hat{M}$ against geographic distance (*d*) should be approximately linear and a significant regression between $\hat{M}$ and *d* should be detectable by a significant regression coefficient. The use of $\hat{M}$ as an estimator of gene flow has advantages over comparing Nei's (1972) genetic distance with geographic distance. Slatkin's $\hat{M}$ is nearly independent of the mutation rate when the mutation rate is small. Because mutation rates can vary from locus to locus, averaging across loci therefore does not confound potentially different mutation rates. $\hat{M}$ also has a greater resolving power because it uses an expanded scale of measurement. At small genetic distances, differences between populations can be slight, whereas values of $\hat{M}$ can be more variable over a larger range. Thus, Slatkin (1993) suggested that $\hat{M}$ is more appropriate for characterizing gene dispersal patterns within a species, whereas genetic distance may be more appropriate for measuring divergence between species. Simulations of a variety of models (Slatkin 1993) indicate that a clear pattern of dependence of $\hat{M}$ on geographic distance can be detected in non-equilibrium populations when at least 10 polymorphic loci are sampled.

This is an important point, considering that populations of lodgepole pine reached their northern limit probably within the last 100 years (Cwynar and MacDonald 1987) and are likely not yet at equilibrium. Although a pattern of isolation by distance may be detected using $\hat{M}$ in non-equilibrium populations, Slatkin (personal communication) indicated that if a pattern is detected, then populations should be close to equilibrium.

Random amplified polymorphic DNA (RAPD) markers are well suited for population genetic analysis because they are primarily neutral and highly variable. Each random primer typically yields several variable RAPDs, and since many random primers are available, a large number of RAPDs are available for analysis. RAPDs have been shown to be inherited in a biparental dominant Mendelian manner (Carlson *et al.* 1991; Roy *et al.* 1992; Heun and Helentjaris 1993; Li and Yeh 2001). Their use as markers for the genetic characterization of populations has been well established for a variety of organisms (Chalmers *et al.* 1992; Huff *et al.* 1993; Yeh *et al.* 1995; Caccone *et al.* 1997; Hogbin *et al.* 1997; Esquibet *et al.* 1998; Lou *et al.* 1998; Suazo *et al.* 1998).

In this chapter the hypotheses of isolation by distance and presence of a disjunct northern refugia are explored in *Pinus contorta* using RAPD markers, with the objective of determining whether a pattern of isolation by distance exists. Subspecies *latifolia* and *contorta* will be the primary focus as seed material was available from only a small number of populations of subspecies *murrayana* and *bolanderi*. Isolation of populations from one another over a long period of time would result in genetic divergence due to the effects of random drift and mutation, regardless of geographic distance. The differentiation between these populations would disrupt any gradual pattern of isolation by distance (by analysis of $\hat{M}$) that would otherwise exist in the absence of isolation. Pairwise comparisons of $\hat{M}$ between isolated populations would yield lower values of $\hat{M}$ than pairwise comparisons between populations that were not isolated from one another.

Materials and methods

Population sampling

Seeds were collected from 31 lodgepole pine populations, spanning most of the range of the species and including all four subspecies (Figure 3-1). The seed collection can be divided into two categories: populations in which the seeds have been bulked, and populations in which lots from individual trees were kept distinct. For populations where seed lot was preserved (16 populations of subspecies *latifolia*), eight to twelve trees per population were sampled, with six seeds analysed from each tree. Thirty seeds per population were used from the bulked samples. Samples from subspecies *contorta*, *murrayana*, and *bolanderi* were all bulked, as were three of the 19 populations from subspecies *latifolia*. The organization of the groups, with the number of seeds used for each is given in Table 3-1. Cones were air dried for up to six months, after which they were opened by being heated in a kiln for 12 hours at 57° C. After being removed from the cones and cleaned, the seeds were stored at –20° C.

DNA extraction and amplification

Prior to DNA extraction, the seeds were re-hydrated overnight in distilled water. Seed coats were removed, and the haploid megagametophyte was separated from the embryo. Total DNA was extracted from each megagametophyte using a phenol/chloroform method modified from that of Lee and Taylor (1990). DNA concentration was quantified on a 0.7% agarose gel by comparison with a λ-DNA standard (Roche). Working samples were diluted to 0.2 – 0.4 ng/µl. PCR amplification of DNA was carried out in a GeneAmp 9600 Thermo-cycler (Perkin-Elmer). Amplification reactions contained: 10 mM reaction buffer (Perkin-Elmer), 0.2 mM each of dATP, dCTP, dGTP, and dTTP, 2.5 mM $MgCl_2$, 0.2 µM primer, 0.6 units of *Taq* DNA polymerase (Perkin-Elmer), 2-4 ng sample DNA, in a total volume of 20 µl. The PCR reaction temperature profile was as follows: 92° C for 2 min.; 40 cycles of 92° C for 30s, 36° C for 30s, 72° C for 1 min., followed by 72° C for 8 min. Amplification products were loaded into a 1.4% agarose gel and electrophoresis was performed at 70V for 5 hours. Gels were stained with ethidium bromide and photographed under UV illumination.

The molecular sizes of RAPD fragments were estimated by comparison with DNA molecular weight marker VI (Roche) using DNA ProScan program (1997). Samples that had amplification failures or unscorable fragments were treated as missing data.

There has been some criticism of the use of RAPDs because they are sensitive to experimental conditions (Smith *et al.* 1994; Harris 1995). We have, however, found that exercising careful control to use a consistent quantity of DNA per amplification (2-4 ng) and using high quality *Taq* Polymerase has resulted in clear, reproducible RAPD profiles. Primer screening is also essential to obtain reproducible RAPD profiles. Two hundred ten-nucleotide random primers from the University of British Columbia Biotechnology Laboratory were initially screened on six megagametophytes from each of five populations. Ten primers that consistently revealed sharp and reproducible RAPDs over several independent runs were selected for this population study.

Data Analysis

The RAPDs were scored as present (1) or absent (0) for individual megagametophytes. RAPD bands are denoted here by the primer followed by a letter, starting with the smallest scorable fragment. For example, fragment 211-A refers to the smallest scorable fragment generated by primer 211. For populations where cone lots were kept separate the maternal genotype was estimated from the six megagametophytes analyzed per tree.

RAPD frequencies were derived for each of the populations. Calculations of allele frequencies were based on the genotype of the maternal tree for populations where this information was known (most populations of subspecies *latifolia*). Estimates of genetic diversity, unbiased genetic distance, and G_{ST} (Nei 1972; Nei 1973; Nei 1978) were calculated for all 31 populations and for each of the subspecies groups.

A comparison of the variance in G_{ST} was performed within and between three subspecies of *Pinus contorta*. This method has been applied previously in *Pinus ponderosa* (Latta and Mitton 1999) using F_{ST}. This analysis is based on probabilities of drifting allele frequencies in structured populations (Robertson 1975). Restricted gene flow between two groups of populations allows allele frequencies to drift independently in each of the groups. This may result in differentiation of the two groups, as detected by G_{ST} (or F_{ST}). However, this result is not guaranteed; allele frequencies at a particular locus may drift in the same direction by chance. If the number of populations is large in each group, the probability of the same allele drifting to high frequency in all populations is small. This outcome is also dependent on the starting allele frequencies. Alleles with high initial frequency have a greater likelihood of remaining at high frequencies in all populations. If the subspecies of *Pinus contorta* have been separated historically, as suggested by Critchfield (1980, 1985), and Wheeler and Guries (1982), the variance of G_{ST} across loci within each of the subspecies should be smaller than the variance of G_{ST} between subspecies. Following Latta and Mitton (1999), values of G_{ST} were arcsine square-root transformed before calculating the variance, because G_{ST} is a proportion.

Phylogenetic trees were constructed using the neighbour-joining method implemented in the program Neighbor in the PHYLIP package of programs (Felsenstein 1993). These trees are based on Nei's pairwise genetic distance. Trees were constructed for all populations and for each of subspecies *latifolia*, *murrayana*, and *contorta*.

TREEVIEW 1.6 (Page 1996) was used to create the displays for figures 3-2, 3-3, 3-4 and 3-5. Subspecies *bolanderi* was grouped with subspecies *contorta* in this analysis because seeds from only one population of subspecies *bolanderi* were available, and it is recently derived from populations of subspecies *contorta* (Aitken and Libby 1994). The Ewens-Watterson test for neutrality (Manley 1985) was performed using POPGENE 1.32 (Yeh *et al.* 1999) to determine whether the RAPDs selected for study did behave as neutral markers. This test was run over all populations as well as separately for each of the subspecies, and the diploid data set. Slatkin's index of genetic similarity ($\hat{M}$) was used to determine gene flow between populations using his isolation by distance program (Slatkin 1993).

$\hat{M}$ is derived from the relationship $F_{ST} = 1/(4M+1)$ where F_{ST} is the genetic divergence between two sampled populations. $\hat{M}$ can be calculated using two different measures of genetic divergence. In a previous study on lodgepole pine using isozymes, Yang and Yeh (1995) employed Weir and Cockerham's $\hat{\theta}$ as an estimator of F_{ST} because it is unbiased (Weir and Cockerham 1984). Using $\hat{\theta}$ however can present problems as it can have a negative value. Alternately, G_{ST} (Nei 1973) can be used to estimate F_{ST}. Although it is more biased than $\hat{\theta}$ there is no problem associated with negative values of genetic divergence. The results from both methods of estimating F_{ST} are presented. Geographic distance (d) between pairs of populations was calculated using ArcView GIS 3.2. A distance finding script using points based on population latitude and longitude was run in projection UTM-1927 (using zone 10 reference). A linear relationship between $\log(\hat{M})$ and $\log(d)$: $\log(\hat{M}) = a + b\log(d)$ is expected if there is a pattern of isolation by distance (Slatkin 1993). A significant regression with negative slope ($b<0$) would indicate such a pattern. The estimates of $\hat{M}$ between population pairs are not independent because each population is involved in multiple estimates. Thus there is no parametric test for significance of b. An empirical distribution of b was generated by bootstrapping following the method of Yang and Yeh (1995). A 95% confidence interval for b was defined as the interval between the 26th and 975th of 1000 ordered bootstrap estimates of b. A significant regression would be one in which the interval excluded zero.

Results

The ten primers used in this study generated a total of 52 RAPD markers used for analysis. Over all populations, 36 of the 52 RAPDs (69%) were polymorphic, ranging from 14 (population 136) to 24 (population 31) polymorphic RAPDs per population. Three RAPDs (243-D, 250-C, and 428-E) were present at a frequency of 0.100 or less in some populations. In six cases a rare RAPD was restricted to a single population: 250-A in population 31, 280-E and 635-E in population 49, 429-E and 429-G in population 33, and 623-B in population 18.

Over all 31 populations, average genetic diversity (Nei's 1978 unbiased h) was 0.1332 (Table 3-2). There were substantial differences between diversity estimates of the four subspecies. Subspecies *latifolia* had the highest mean heterozygosity (0.1480) followed by subspecies *murrayana* (0.1147) and *contorta* (0.1111). Subspecies *bolanderi* had the lowest value (0.0791).

Average genetic distance within and between subspecies is presented in (Table 3-3). Genetic distances between populations within subspecies *latifolia* and *murrayana* were similar and both had a similar level of genetic distance from subspecies *contorta*. The highest values of genetic distance were between subspecies *bolanderi* and all other subspecies, indicating it is genetically the most unique of the subspecies. Of the four subspecies, *contorta* has the highest within-subspecies genetic distance and is quite distinct from *bolanderi* despite their close geographical proximity and the supposed recent origin of subspecies *bolanderi* from subspecies *contorta*.

The degree of differentiation between groups was determined by estimation of G_{ST} (Nei 1973). G_{ST} was defined by Nei (1973) as a ratio of the average gene diversity between subpopulations (D_{ST}) to the gene diversity in the total population (H_T). G_{ST} is equal to the weighted average of F_{ST} (Nei 1973). Values of G_{ST} for subspecies *latifolia*, *murrayana*, and *contorta* are 0.067, 0.036, and 0.079 respectively. The overall value for G_{ST} across all populations is 0.090.

The transformed variances of the G_{ST} values for subsp. *latifolia*, *murrayana*, and *contorta* were 0.028, 0.024, and 0.049 respectively. The value for G_{ST} between these three subspecies was 0.014 with a transformed variance of 0.007.

The phylogenetic tree constructed from all 31 populations is presented in Figure 3-2. Although some patterns in the relationship between populations and geographic proximity can be seen (e.g., populations of subspecies *latifolia* from the southern part of its range share a common node), no clear division of the four subspecies into distinct clades is seen. When populations of just subspecies *latifolia* are examined (Figure 3-3), two major groups are seen that correspond to a general north-south trend. The tree generated from subspecies *contorta* and *bolanderi* however did not display such a pattern (Figure 3-4). Similarly, the populations of subspecies *murrayana* did not show any grouping of populations that were geographically close together (Figure 3-5), however this could be due to the small number of populations analyzed in this subspecies.

The results of the Ewens-Watterson test for neutrality are displayed in Table 3-4. The number of non-neutral loci ranges from 1 (subspecies *contorta* and *bolanderi*) to 5 (subspecies *murrayana*). Overall, the vast majority of RAPD markers (94%) are neutral, as would be expected from using random primers.

The pairwise values of genetic similarity ($\hat{M}$) are plotted against geographic distance (d) on a logarithmic scale in Figure 3-6. Graphs for all 31 populations as well as subsets consisting of the three subspecies *latifolia*, *contorta* and *murrayana* are presented for $\hat{M}$ calculated using both $\hat{\theta}$ and G_{ST} as an estimator of F_{ST}. Genetic similarity for population pairs was quite variable, ranging from –32.55 to 207.63 using $\hat{\theta}$ and 0.84 to 14.1 using G_{ST}. Two population pairs (57 and 16, 69 and 16) had negative values for $\hat{M}$ when calculated from $\hat{\theta}$. It can be seen from the equation $F_{ST} = 1/(4M+1)$ that the estimator of F_{ST}, ($\hat{\theta}$) is negative for these two pairs. According to Slatkin (1989), this must be interpreted as meaning that F_{ST} is so small that significant population differences cannot be detected. A few populations have $\hat{M}$ values at the other extreme. Eighteen of 465 population pairs based on $\hat{\theta}$ and four based on G_{ST} have $\hat{M}$ values greater than 10. Values that are this high can be taken to mean effective panmixia (Slatkin 1993). The dotted lines in Figure 3-6 (at $\hat{M} = 1$) indicate the point at which the relative effects of gene flow or random drift are predominant (Wright 1943). Below this line, a lack of gene flow between the pairs results in population differentiation due to drift. As can be seen from Figure 3-6 (a), a number of population pairs have values of $\hat{M}$ that are less than 1. The majority of these are between populations of different subspecies. Within subspecies *contorta* however (Figure 3-6 (e)), more than half of the $\hat{M}$ values based on $\hat{\theta}$ are below 1. The mean values of $\hat{M}$ and *d* for each of the four groups are presented in Table 3-5.

The regressions of log($\hat{M}$) on log(*d*) were negative for each of the subspecies considered, and for the group of all 31 populations. As shown in Table 3-6, the confidence intervals generated by bootstrapping indicate that *b* is significantly different from zero for all populations taken together, and for subspecies *latifolia*. This indicates that a pattern of isolation by distance is present in *latifolia* and the species as a whole.

Discussion

The levels of diversity estimated here from RAPD markers (Table 3-2) are quite similar to other reports for *Pinus contorta*. For subspecies *latifolia*, estimates of *h* based on isozymes of 0.160 (Yeh and Layton 1979), 0.118 (Wheeler and Guries 1982), 0.184 (Dancik and Yeh 1983) and 0.194 (Yang and Yeh 1995) are comparable to the value obtained in this study (0.151). Using RAPDs, Ye *et al*. (2002) and Thomas *et al*. (1999) reported an overall *h* for *Pinus contorta* subsp. *latifolia* of 0.166 and 0.43 respectively. For subspecies *contorta*, and *murrayana* Wheeler and Guries (1982) report average *h* (from isozymes) of 0.126 and 0.124 respectively, as compared to Yang and Yeh (1995) (from isozymes) 0.181, 0.196 and this study 0.113 and 0.116. The valued obtained here for *h* (0.081) from subspecies *bolanderi* is also close to the value obtained by Wheeler and Guries (1982) ($h = 0.109$) and Aitken and Libby (1994) ($h = 0.096$) from isozyme data. Aitken and Libby (1994) showed that subspecies *bolanderi* was derived from populations of subspecies *contorta*. The small level of diversity in this subspecies may be a result of founder effects and continued small population sizes.

While the values for h obtained from different marker types are alike, it can be seen that estimates based on isozymes are generally higher than those obtained from RAPD markers. This may be due to the dominant nature of RAPD markers. Since there are only two possible allele states, the maximum number of mean alleles per locus is 2. In this unlikely case, all loci would be polymorphic. A more reasonable percentage of polymorphic loci such as 40% (the range seen here is 37% – 46%) yields a mean number of alleles per locus of 1.40. In contrast, isozymes can potentially have a much greater number of alleles per locus (many theories assume an infinite alleles model), which could contribute to a higher measure of gene diversity.

Polymorphisms in RAPD markers are thought to be primarily due to differences in nucleotides at primer annealing sites between homologous regions of DNA (Williams *et al.* 1990). The majority of nucleotide changes will be either selectively neutral, or negative. Negative changes will be quickly eliminated from the population. Selectively neutral changes however may persist in a population and increase in frequency. Such neutral markers are optimal for characterizing relationships within and between populations because their distributions are not confounded by selection. The results of the Ewens-Watterson test of neutrality confirm the expectation that RAPD markers are primarily selectively neutral. Across all populations, only three loci (211-D, 250-A, 250-C) exceeded the 95% confidence intervals, failing the test of neutrality.

The phylogenetic tree generated by the neighbour-joining procedure failed to clearly group the four subspecies analyzed (Figure 3-2). This indicates that the subspecies of lodgepole pine have not been isolated from each other long enough for them to clearly diverge genetically, or that there has been sufficient gene flow between the subspecies to prevent differentiation. UPGMA analysis of isozyme frequencies of lodgepole pine by Wheeler and Guries (1982) was also unable to differentiate subspecies.

The analysis of the variance in G_{ST} also indicates that the allele frequencies of the subspecies have not drifted apart. Much more heterogeneity in values of G_{ST} is seen within each of the subspecies than is seen between them. This is what would be expected if the separation of the subspecies were too recent to be detected with RAPD markers.

These two results suggest that the subspecies designations in common use are not supported by the genetic relationships between populations, and indicates that prior to post-glacial expansion, populations of *Pinus contorta* may have existed in a panmictic state.

Almost all of Canada was covered with ice during the last glaciation (Dyke and Prest 1987). Most species now present in this area expanded into their current range from unglaciated areas south of the ice as it retreated, opening land for colonization. An ice-free area also occurred northwest of the Cordilleran ice sheet in areas of Yukon and Alaska, which served as a refugium for many arctic plant species (Hultén 1937). It has also been proposed that the Pacific Northwest had ice-free areas along the coastline that may have served as refugia (Heusser 1960, Terasmae 1973). Although the Cordilleran ice sheet may have extended right to the ocean over much of the coast, there may have been localized unglaciated areas (Dahl 1946). The coastal shoreline was certainly climatically different at the time of the glacial maximum and also quite distant from its current location, as a result of lower ocean levels. Maps of the extent of the Wisconsin glaciation provided by Dyke and Prest (1987) indicate that the ocean level was approximately 100 m lower than it is today. Significantly more land mass would have been exposed at 18000 to 14000 years before present. Vancouver Island would have been part of the mainland, and the Hecate Strait between the Queen Charlotte Islands and the mainland might have been only a few kilometres across at some points (Dyke and Prest 1987).

The history of lodgepole pine and the possibility of its survival in such unglaciated areas have been discussed in a number of previous studies (Hanson 1950; Heusser 1967; Anderson 1970; Terasmae 1973; von Rudloff and Nyland 1979; Critchfield 1980, 1985; Wheeler and Guries 1982; MacDonald and Cwynar 1985, 1991; Cwynar and MacDonald 1987; Yang and Yeh 1995). Inferences of the post-glacial history from genetic relationships between populations however have been at odds with data obtained from palynological studies. While unique morphological and genetic characteristics of populations in the Yukon have led some (e.g. von Rudloff and Nyland 1979; Wheeler and Guries 1982) to propose that *Pinus contorta* subsp. *latifolia* persisted in this region during the Wisconsin glaciation, data from fossil pollen study (MacDonald and Cwynar 1985) indicates a much more recent arrival of pine in the Yukon.

Wright (1943) originally proposed that the amount of gene flow between populations within a species should be a function of distance. This idea can be used to evaluate whether the current distribution of the various subspecies of lodgepole pine is from one or multiple refugia. If populations had persisted in northern or coastal refugia throughout the Wisconsin glaciation (approximately 100,000 years) in numbers so small as to not contribute to pollen deposition, the effects of random drift in isolation from populations south of the ice sheet ought to have resulted in significant divergence from populations located south of the ice margin. Any trend of isolation by distance, or relationship between genetic similarity and geographic distance would be disrupted (Yang and Yeh 1995).

The two estimators of F_{ST} used here each have relative merits and weaknesses. $\hat{\theta}$ is intended to be an unbiased estimator (Weir and Cockerham 1984), although it can have negative values, as was seen between two of the population pairs. G_{ST} does not consider the number of locations or number of individuals sampled, however it will always be >0. In simulation studies, Slatkin and Barton (1989) found that at high levels of gene flow, $\hat{\theta}$ tends to overestimate $\hat{M}$ whereas G_{ST} tends to underestimate it. As can be seen in Figure 3-6 (e) and Figure 3-6 (f), values of $\hat{M}$ are very similar between the two estimators at moderate levels of gene flow ($\hat{M} \approx 1$). At such levels, values of $\hat{M}$ based on G_{ST} are roughly twice those based on $\hat{\theta}$, as is expected from the equations (Slatkin and Barton 1989). The discussion of the results for the estimates of slope (b) from the regression of log ($\hat{M}$) on log (d) will be treated separately for subspecies *latifolia*, *contorta* and *murrayana*. The representation of subspecies *bolanderi* by a single population prevents such analysis.

Subspecies *latifolia*

The ice-free area northwest of the Cordilleran ice sheet in portions of the Yukon and Alaska was hypothesized to have been a refugium for *Pinus contorta* subsp. *latifolia*, from which present day populations in the Yukon and northern British Columbia were derived (Hanson 1950; Anderson 1970; Wheeler and Guries 1982). Using isozyme data, Wheeler and Guries (1982) demonstrated that pine populations in the Yukon region were distinct from other populations of *Pinus contorta* subsp. *latifolia* to the south, had a high number of rare alleles and slightly higher levels of heterozygosity. These authors suggested these characteristics provide a compelling argument that lodgepole pine has expanded south from northern refugia into its current distribution in the Yukon and northern British Columbia, where it would have met pine populations expanding northward from south of the Cordilleran ice sheet.

The concept of a northern refugium for lodgepole pine has been contested by MacDonald and Cwynar (1985) who used palynological data to establish a clear trend of north-westward expansion from south-eastern British Columbia to the limit of the range of lodgepole pine in the Yukon. According to their data, pine reached an area in the Yukon near Buggy Pond approximately 430 years ago. Such hard evidence is difficult to dispute, and MacDonald and Cwynar (1985) assert that the morphological and genetic differences apparent between northern and southern populations of *Pinus contorta* subsp. *latifolia* cannot be explained by invoking a theory of northern refugia.

As can be seen from Table 3-6 the estimate for *b* is negative and significant, indicating that subspecies *latifolia* demonstrates a pattern of isolation by distance. This result agrees with the palaeobotancial interpretation of pollen profiles (MacDonald and Cwynar 1985), but contrasts the conclusions of Wheeler and Guries (1982), and Yang and Yeh (1995), which are based on genetic relationships. Although Wheeler and Guries (1982), found a high number of rare alleles in Yukon and northern British Columbia populations, when their data were reworked by Cwynar and MacDonald (1987) it was found that the mean number of alleles per locus was negatively correlated ($r^2 = -0.468$) with the time since establishment of the population (based on presence of fossil pollen). Thus, the most recently established populations (Yukon and northern British Columbia) are most likely to have a lower number of alleles. As well, Yang and Yeh (1995) found that the number of shared rare alleles in *Pinus contorta* subsp. *latifolia* exhibited a negative correlation with latitude ($r^2 = -0.421$). Both these results are consistent with the expansion of lodgepole pine from refugia south of the Cordilleran ice sheet. A repeated founder effect from the establishment of populations is likely to result in loss of alleles. The third criteria that Wheeler and Guries (1982) use (in a corroborative context) is an expectation of reduced levels of heterozygosity at the edge of range expansion. Their data do not fit this criterion for subsp. *latifolia* and as can be seen in Table 3-2, levels of heterozygosity based on RAPD markers are also not reduced in Yukon or northern British Columbia populations as compared to others within the subspecies. Reduced levels of heterozygosity however are not expected from repeated founding events unless the founding population is small, and increases in number only slowly (Nei *et al.* 1975). Thus, a level of diversity in northern populations that is similar to more southern populations is not sufficient evidence on its own for the presence of a Yukon refugium. The significant pattern of isolation by distance from the regression of log($\hat{M}$) on log (*d*) as determined from RAPD markers suggests that lodgepole pine did not survive the Wisconsin glaciation in ice-free areas of the Yukon.

It is clear however that the Yukon populations do have some unique characteristics. An examination of the RAPD data reveals the existence of unique alleles in some of the northern populations. Population 30 has one locus (250-A) and population 33 has two loci (429-E, 429-G) that are polymorphic and were found in no other population. Additionally, populations 31and 33 share a rare polymorphism (at locus 429-C) found in only two other populations. Of a total of five unique RAPD polymorphisms (present in only one population) in *Pinus contorta* subsp. *latifolia*, three are found in northern populations. The isozyme data of Wheeler and Guries (1982) and Yeh and Layton (1979) both show rare alleles restricted to Yukon populations. A clue to the reason for this may be found in the analysis done by Wheeler and Guries (1982). They noted that a number of the rare alleles present in the Yukon and northern British Columbia are also present in subspecies *contorta*. Although they did not advance the idea, it is probable that there was (and perhaps still is) some level of gene flow occurring between these two subspecies. The most northern coastal population they sampled is close to fiords that cut through the coastal mountains, providing potential routes for migration between the subspecies. If populations of *Pinus contorta* subsp. *contorta* were able to expand up drainages past the end of the fiord, they could potentially come within 10 km of Atlin Lake. Such proximity to populations of *Pinus contorta* subsp. *latifolia*, would easily allow gene flow to occur between the two subspecies. Indeed, three of the four rare alleles found in the most northern coastal population in the study by Wheeler and Guries (1982) are also found in the nearest Yukon population they sampled. The phylogenetic tree generated from all 31 populations (Figure 3-2) also shows that populations of subspecies *contorta* tend to group with populations of subspecies *latifolia*. For example, the most northerly populations of subspecies *contorta* (49 and 50) are closely aligned with the most northerly population of subspecies *latifolia* (33).

It is also interesting to note that data from fossil pollen samples at Waterdevil Pond, British Columbia indicate the establishment of lodgepole pine at 3,190 years before present (MacDonald and Cwynar 1985). This date does not fit the otherwise linear trend of establishment times in the north-westward progression of pine as determined by fossil pollen. Pine appeared at Waterdevil Pond 700 years before the establishment of pine at Kettlehole Pond only 75 km to the northeast. To explain this anomaly, Spear and Cwynar (1997) investigated the possibility of migration of *Pinus contorta* over the White Pass in northern British Columbia. Based on palynological data these authors established that trees and shrubs were not able to use the pass as a migration route from coastal to interior areas. The early presence of pine at Waterdevil pine suggests migration of subsp. *contorta* up a river valley. Forrest (1980, 1981) showed a cline in terpene levels along the Skeena River valley indicating introgression in this valley. The Skeena or other major valley systems, such as the Stikine may have served as migration routes for subsp. *contorta* into the interior before the northwestward advance of subsp. *latifolia*.

The phylogenetic tree generated for subspecies *latifolia* (Figure 3-3) displays two major groups. One group contains populations primarily from southern British Columbia and Alberta, Idaho and Montana. The second group is composed primarily of populations north of this area. A similar pattern of population structure within subspecies *latifolia* was also detected by Wheeler and Guries (1982) and may reflect northward migration from two different southern refugia, located on either side of the Rocky Mountains as suggested by Wheeler and Guries (1982). Alternatively, the pattern may be a result of reduced gene flow between groups subsequent to migration due to the mountain barrier.

Subspecies *contorta*

Hopkins (1972) concluded that subspecies *contorta* expanded into its current range from south of the ice margin. This should result in a pattern of isolation by distance, with populations that are geographically closer together having a greater similarity than populations further apart. The mean value for $\hat{M}$ from subspecies *contorta* was much less than that of subspecies *latifolia*, indicating that coastal populations have less gene flow occurring between populations. In contrast to subspecies *latifolia*, no pattern of isolation by distance was detected. Although b was negative, it was small and not significantly different from zero as judged by the 95% confidence intervals. Slatkin (1993) notes that a pattern of isolation should be easier to detect in populations distributed linearly, which underscores the failure to detect such a pattern in subsp. *contorta*.

Shore pine, as this subspecies is known, has different ecological characteristics and a different life history as compared to subspecies *latifolia* (Critchfield 1980). Cone serotiny is absent or infrequent, so seeds can be released immediately after maturity and have to compete with other species for survival. As a result, populations are less even aged and less densely stocked than those of subspecies *latifolia*. Their reduced effective population size (Yang and Yeh 1995) would allow greater opportunity for local non-adaptive differentiation (Wright 1943). A majority of $\hat{M}$ values <1 (Figure 3-6 (e)) indicates that gene flow between the pairs of populations is not great enough to prevent population differentiation due to drift. Populations of subspecies *contorta* are the most variable among the four subspecies of lodgepole pine. The genetic distance within the subspecies is more than twice that of either subspecies *latifolia* or *murrayana*. As well, there are a number of loci that are polymorphic in only one or two populations within the subspecies. The phylogenetic tree (Figure 3-4) for populations of subspecies *contorta* (and the single population of subspecies *bolanderi*) shows no clear pattern; consistent with populations in which the effect of gene flow is small in relation to genetic drift. The value for G_{ST} is also highest (0. 079) in subspecies *contorta*.

The evidence therefore indicates that it is unlikely that subspecies *contorta* expanded northward from a single refugium south of the glacial margin. This agrees with Wheeler and Critchfield (1982) but contrasts with Yang and Yeh (1995) who found a pattern of isolation by distance in subsp. *contorta*. It should be noted that the pattern found by Yang and Yeh (1995) was weak, and that the majority of their sampling of *Pinus contorta* subsp. *contorta* was south of the Puget Sound area. The palaeobotancial history of subsp. *contorta* also lends support to the idea that coastal populations originated from multiple refugia. Heusser (1960) presented palynological evidence of the presence of lodgepole pine near Juneau, Alaska at a minimum of 10,000 to 11,500 years before present. In order for pine to have reached this location as a result of expansion from south of the ice margin, Heusser (1965) notes that the migration rate for *Pinus* would have to have been approximately 1 km/year. Peteet (1991) found that a similar rate of migration would be required to explain the minimum arrival time of pine at Langara Island (in the Queen Charlottes), which is two to ten times faster than other arboreal coastal species. When this is considered together with the glaciation maps provided by Dyke and Prest (1987) it appears that lodgepole pine took advantage of newly exposed land surface as the ice margin shifted. At 18,000 years ago the Queen Charlotte Islands were covered in ice, while Vancouver Island was mostly ice-free, and connected to the mainland (Figure 3-7). At 14,000 years ago, the ice margin has shifted to cover most of Vancouver Island, but the Queen Charlotte Islands were free of the glaciation and remain so to the present (Figure 3-8). The low ocean levels and greater land surface (as compared to today) meant that the distance between the Queen Charlotte Islands and Vancouver Island was much smaller than it is today. Thus, an opportunity existed for colonization of the Queen Charlotte Islands, after which populations there would have become isolated from those south of the ice.

Peteet (1991) also notes that the fossil pollen record from the Queen Charlotte Islands and areas north indicate the presence of significant amounts of pine pollen in the basal pollen zones. This implies that *Pinus* was present along this entire section of coast even before the formation of the particular lake or bog that was sampled (> 10,500 years before present) (Peteet 1991). It is possible that *Pinus contorta* subsp. *contorta* survived glaciation along portions of the coast as Dahl (1946) suggested. If this were the case however, one might expect much more differentiation between *contorta* and the other subspecies to have developed over the Wisconsin given the duration of separation (~80,000 years) and likely small population sizes. Seeds originating from the Queen

Charlotte Islands may have been able to cross the then narrow Hecate Strait, establishing populations on the coast as the glacier retreated. Peteet (1991) also conjectured that ocean transport of seeds or cones might have facilitated the dispersal of subsp. *contorta*. While some expansion from populations in the Puget sound area is likely, as suggested by Hopkins (1972), these populations were probably not the only source for recolonization.

Subspecies *murrayana*

The degree of genetic similarity (mean $\hat{M} = 2.72$) among populations of subspecies *murrayana* is greater than that of subspecies *contorta*, but still considerably less than that of subspecies *latifolia*. The regression of $\log(\hat{M})$ on $\log(d)$ yielded a value for b of -0.33746. This result however is not significant, as the confidence intervals encompass zero. Thus, subspecies *murrayana* shows no pattern of isolation by distance. The confidence intervals are also very large, which is probably a function of the small sample size for this subspecies; a greater sample size may reveal that such a pattern exists. The current range of subspecies *murrayana* is located in areas that were south of the Wisconsin maximum and little information is available about its distribution during that time. It is unclear whether subspecies *murrayana* and *latifolia* were physically isolated during this time or whether subspecies *latifolia* was even a separate entity. The genetic distance within each of the two subspecies is similar and only slightly smaller than the distance between them. It is possible that cone serotiny which is present (although not widespread) in subspecies *murrayana* proved to be an important adaptive trait as populations expanded northward. The effects of selection associated with range expansion, along with repeated founding events, may have been the primary causes of differences between the subspecies observed today.

Conclusion

The overall level of differentiation at RAPD loci between populations of *Pinus contorta* is similar to what has been observed in other conifers (Hamrick *et al*. 1992), which is consistent with the typically extensive amounts of gene flow observed in this group (Govindaraju 1988). However the pattern of differentiation is different between the subspecies of *Pinus contorta*.

For subspecies *latifolia*, geographically close populations usually have allele frequencies that are similar, whereas geographically distant populations are more divergent. This pattern of isolation by distance is what would be expected if populations originated from a single refugium south of the ice and expanded north following glacial retreat.

This contrasts with the lack of such a pattern in subspecies *contorta*. Differentiation and genetic distance between populations is considerably greater than in the other subspecies, suggesting that populations did not migrate in a stepwise fashion along the coast. If this were the mode of migration, geographically close populations would be expected to have a greater genetic affinity than geographically distant populations.

The small number of populations sampled in subspecies *murrayana* (and the resulting wide bootstrapped error for *b*) prevents definitive conclusions with respect to a pattern of isolation by distance. The values of genetic distance between subspecies *murrayana* and *latifolia* however suggest a close relationship. The current restricted opportunity for gene flow (in the Cascades region) between these subspecies indicates that populations may have been in continuous contact during the Wisconsin glaciation.

The data presented here fit well with what is known about the postglacial history of lodgepole pine as evidenced from palynological research. The longstanding debate over the persistence of *Pinus contorta* subsp. *latifolia* in a Yukon refugium during the Wisconsin glaciation has centred on the unique characteristics of populations in this region. The evidence here suggests that *Pinus contorta* subsp. *latifolia* did not persist in the Yukon. The unique morphological and genetic characteristics are most likely a result of gene flow from nearby populations of subspecies *contorta* in this area. It is unlikely that northward migration from populations surviving south of the glacial margin was the primary mode of colonization for subspecies *contorta*. Subspecies *contorta* has probably colonized much of its northern range from the Queen Charlotte Islands, with potentially additional subsequent migration from populations surviving south of the ice margin.

References

Aitken, S.N. and W.J. Libby. 1994. Evolution of the pygmy-forest edaphic subspecies of *Pinus contorta* across an ecological staircase. Evolution **48**: 1009-1019.

Baker, R.G. 1970. Pollen sequence from the late Quaternary sediments in Yellowstone Park. Science **168**: 1449-1450.

Caccone, A., G. Allegrucci, C. Fortunato and V. Sbordoni. 1997. Genetic differentiation within the European sea bass (*D. labrax*) as revealed by RAPD-PCR assays. J. Hered. **88**: 316-324.

Calder, J. and R. Taylor. 1968. Flora of the Queen Charlotte Islands. Can. Dep. Agric. Res. Branch Monogr. No. 4, Part 1.

Carlson, J.E., L.K. Tulsieram, J.C. Glaubitz, V.W.K. Luk, C. Kauffeldt and R. Rutledge. 1991. Segregation of random amplified DNA markers in F_1 progeny of conifers. Theor. Appl. Genet. **83**: 194-200.

Carrara, P.E. and R.E. Wilcox. 1984. Glacial chronology during the last 12,000 years in the Glacier National Park region, Montana. Abstracts of the Eighth Biennial Meeting, American Quaternary Association, University of Colorado, Boulder, Co, August 1984.

Chalmers, K.J., R. Waugh, J.I. Sprent, A.J. Simons and W. Powell. 1992. Detection of genetic variation between and within populations of *Gliricidia sepium* and *G. maculata* using RAPD markers. Heredity **69**: 465-472.

Critchfield, W.B. 1980. Genetics of Lodgepole Pine. U.S. For. Serv. Res. Pap. WO-37.

Critchfield, W.B. 1985. The late Quaternary history of lodgepole and jack pines. Can J. For. Res. **15**: 749-772.

Cwynar, L.C. and G.M. MacDonald. 1987. Geographical variation of lodgepole pine in relation to population history. Am. Nat. **129**: 463-469.

DNA Proscan Inc. 1997. Pro-RFLP Version 2.34. Nashville, Tennessee.

Dancik, B.P. and F.C. Yeh. 1983. Allozyme variability and evolution of lodgepole pine (*Pinus contorta* var. *latifolia*) and jack pine (*P. banksiana*) in Alberta. Can. J. Genet. Cytol. **25**: 57-64.

Dyke, A.S. and V.K. Prest. 1987. Late Wisconsin and Holocene history of the Laurentide ice sheet. Geog. Phys. Quat. **41**: 237-263.

Esquibet, M., S. Bekal, P. Castagnone-Sereno, J.P. Gauthier, R. Rivoal and G. Caubel. 1998. Differentiation of normal and giant *Vicia faba* populations of the stem nematode *Ditylenchus dipsaci*: agreement between RAPD and phenotypic characteristics. Heredity **81**: 291-298.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Forrest, G.I. 1980. Geographical variation in the monoterpenes of *Pinus contorta* oleoresin. Biochem. Syst. Ecol. **8**: 343-359.

Forrest, G.I. 1981. Geographical variation in oleoresin monoterpene composition of *Pinus contorta* from natural stands and planted seed collections. Biochem. Syst. Ecol. **9**: 97-103.

Govindaraju, D.R. 1988. Relationship between dispersal ability and levels of gene flow in plants. Oikos **52**: 31-35.

Hamrick, J.L., M.J.W. Godt and S.L. Sherman-Broyles. 1992. Factors influencing levels of genetic diversity in woody plant species. New Forests **6**: 95-124.

Hansen, H.P. 1950. Postglacial forests along the Alaska highway in British Columbia. Proc. Am. Philos. Soc. **94**: 411-422.

Harris, S.A. 1995. Systematics and randomly amplified polymorphic DNA in the genus *Leucaena* (Leguminosae, Mimosoideae). Pl. Syst. Evol. **197**: 195-208.

Heun, M. and T. Helentjaris. 1993. Inheritance of RAPDs in F_1 hybrids of corn. Theor. Appl. Genet. **85**: 961-968

Heusser, C.J. 1960. Late Pleistocene environments of north Pacific North America. Am. Geogr. Soc. Special Publ. No. 35.

Hogbin, P.M., D.J. Ayre and R.J. Whelan. 1998. Genetic variation and reproductive success of road verge populations of the rare shrub *Grevilla barklyana* (Proteaceae). Heredity **80**: 180-186.

Huff, D.R., R. Peakall and P.E. Smouse. 1993. RAPD variation within and among natural populations of outcrossing buffalograss [*Buchloe dactyloides* (Nutt.) Engelm.]. Theor. Appl. Genet. **86**: 927-934.

Hultén, E. 1937. Outline of the history of arctic and boreal biota during the Quaternary period. Bokforlags Aktiebolaget Thule, Stockholm, Sweden.

Latta, R.G and J.B. Mitton. 1999. Historical separation and present gene flow through a zone of secondary contact in Ponderosa pine. Evolution **53**: 769-776.

Lee, S.B. and J.W. Taylor. 1990. Isolation of DNA from Fungal Mycelia and Single Spores. *In*: PCR Protocols. Edited by M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White. Academic Press Inc., San Diego, Cal. pp. 282-287.

Li, C., and F.C. Yeh. 2001. Construction of a framework map in *Pinus contorta* subsp. *latifolia* using random amplified polymorphic DNA markers. Genome **44**: 147-153.

Lou, K.F., M.J. Weiss, P.L. Bruckner, W.L. Morrill, L.E. Talbert and J.M. Martin. 1998. RAPD variation within and among geographic populations of wheat stem sawfly (*Cephus cinctus* Norton). J. Hered. **89**: 329-335.

MacDonald, G.M. and L.C. Cwynar. 1985. A fossil pollen based reconstruction of the late Quaternary history of lodgepole pine (*Pinus contorta* ssp. *latifolia*) in the western interior of Cananda. Can. J. For. Res. **15**: 1039-1044.

MacDonald, G.M. and L.C. Cwynar. 1991. Post-glacial population growth rates of *Pinus contorta* ssp. *latifolia* in western Canada. J. Ecol. **79**: 417-429.

Mack, R.N., N.W. Rutter, V.M. Bryant Jr. and S. Valastro. 1978. Late Quaternary pollen record from Big Meadow, Pend Oreille County, Washington. Ecology **59**: 956-966.

Manly B.F.J. 1985. The Statistics of Natural Selection. Chapman and Hall, London.

Nei, M. 1972. Genetic distance between populations. Am. Nat. **106**: 283-292.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Nat. Acad. Sci. **70**: 3321-3323.

Nei, M., T. Maruyama and R. Chakraborty. 1975. The bottleneck effect and genetic variability in populations. Evolution **29**: 1-10.

Nei, M. 1978. Estimation of average heterzygosity and genetic distance from a small number of individuals. Genetics **89**: 583-590.

Page, R.D.M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Comp. Appl. Biosci. **12**: 357-358.

Peteet, D.M. 1991. Postglacial migration history of lodgepole pine near Yakutat, Alaska. Can. J. Bot. **69**: 786-796.

Robertson, A. 1975. Gene frequency distributions as a test of selective neutrality. Genetics **81**: 775-785.

Roy, A., Frascaria, N., Mackay, J., and Bousquet, J. 1992. Segregating random amplified polymorphic DNAs (RAPDs) in *Betula alleghaniensis*. Theor. Appl. Genet. **85**: 173-180.

Slatkin, M. and N.H. Barton. 1989. A comparison of three indirect methods for estimating average levels of gene flow. Evolution **43**: 1349-1368.

Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution **47**: 264-279.

Smith, J.J., J.S. Scott-Craig, J.R. Leadbetter, G.L. Bush, D.L. Roberts and D.W. Fulbright. 1994. Characterization of random amplified polymorphic DNA (RAPD) products from *Xanthomonas campestris* and some comments on the use of RAPD products in phylogenetic analysis. Mol. Phyl. Evol. **3**: 135-145.

Spear, R.W. and L.C. Cwynar. 1997. Late Quaternary vegetation history of White Pass, northern British Columbia, Canada. Arc. and Alp. Res. **29**: 45-52.

Suazo, A., R. McTiernan and H.G. Hall. 1998. Differences between African and European honey bees (*Apis mellifera* L.) in random amplified polymorphic DNA (RAPD). J. Hered. **89**: 32-36.

Thomas, B.R., S.E. Macdonald, M. Hicks, D.L. Adams and R.B. Hodgetts. 1999. Effects of reforestation methods on genetic diversity of lodgepole pine: an assessment using microsatellite and randomly amplified polymorphic DNA markers. Theor. Appl. Genet. **98**: 793-801.

von Rudloff, E. and E. Nyland. 1979. Chemosystematic studies in the genus *Pinus*. III. The leaf oil terpene composition of lodgepole pine from the Yukon Territory. Can. J. Bot. **57**: 1367-1370.

Weir, B.S. and C.C. Cockerham. 1984. Estimating *F*-statistics for the analysis of population structure. Evolution **38**: 1358-1370.

Wheeler, N.C., and R.P. Guries. 1982. Biogeography of lodgepole pine. Can J. Bot. **60**: 1805-1814.

Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Acids Res. **18**: 6531-6535.

Wright, S. 1943. Isolation by distance. Genetics **28**: 114-118.

Yang, R.-C. and F.C. Yeh. 1995. Patterns of gene flow and geographic structure in *Pinus contorta* Dougl. For. Genet. **2**: 65-75.

Yeh, F.C. and C. Layton. 1979. The organization of genetic variability in central and marginal populations of lodgepole pine *Pinus contorta* spp. *latifolia*. Can. J. Gen. Cyt. **21**: 487-503.

Yeh, F.C., D.K.X. Chong, and R.-C. Yang. 1995. RAPD variation within and among natural populations of trembling aspen (*Populus tremuloides* Michx.) from Alberta. J. Hered. **86**: 454-460.

Yeh, F.C., R.-C. Yang, J. Mao, Z. Ye and T.J.B. Boyle. 1996. POPGENE, the Microsoft Windows-based user-friendly software for population genetic analysis of co-dominant and dominant markers and quantitative traits. Department of Renewable Resources, University of Alberta, Canada.

Table 3-1: Seed organization and number of seeds analyzed for the 31 populations of *Pinus contorta* used in this study.

Population	Subspecies	Category	No. of megagametophytes sampled	No. of maternal trees sampled
34	*latifolia*	Not bulked	66	11
35			66	11
68			72	12
69			72	12
31			48	8
30			54	9
67			72	11
36			54	9
28			48	8
27			72	12
60			54	9
61			72	12
18			60	10
57			48	8
16			72	12
33			72	12
144		Bulked	30	
145			30	
146			30	
84	*contorta*		30	
130			30	
128			30	
118			30	
112			30	
50			30	
49			30	
119	*murrayana*		30	
127			30	
133			30	
135			30	
136	*bolanderi*		30	

Table 3-2: Observed and estimated heterozygosity values for 31 populations of lodgepole pine.

	Population	Expected *h* [†]	Unbiased Expected *h* [‡]	Observed *h*
latifolia	34	0.1464	0.1539	0.1635
	35	0.1418	0.1393	0.1521
	68	0.1390	0.1084	0.1493
	69	0.1446	0.1426	0.1554
	31	0.1582	0.1592	0.1707
	30	0.1453	0.1443	0.1627
	67	0.1500	0.1466	0.1667
	36	0.1607	0.1605	0.1752
	28	0.1706	0.1719	0.1926
	27	0.1548	0.1601	0.1715
	60	0.1579	0.1617	0.1931
	61	0.1698	0.1433	0.2035
	18	0.1565	0.1572	0.1766
	57	0.1609	0.1620	0.1803
	16	0.1530	0.1520	0.1811
	33	0.1512	0.1543	0.1631
Mean		0.1538	0.1511	0.1723
	144	0.1579	0.1535	
	145	0.1333	0.1288	
	146	0.1106	0.1066	
Mean		0.1510	0.1480[a]	
contorta	84	0.0987	0.0938	
	130	0.1254	0.1201	
	128	0.1277	0.1287	
	118	0.0987	0.0975	
	112	0.0934	0.0903	
	50	0.1079	0.1082	
	49	0.1396	0.1393	
Mean		0.1131	0.1111[b]	
murrayana	119	0.1334	0.1345	
	127	0.1139	0.1172	
	133	0.1108	0.1055	
	135	0.1072	0.1016	
Mean		0.1163	0.1147[c]	
bolanderi	136	0.0809	0.0791	
Grand Mean		0.1357	0.1332	

[†] Expected heterozygosity from Nei 1973. [‡] Unbiased expected heterozygosity from Nei (1978). *a* significantly different from *b* and *c* at $p < 0.01$.

Table 3-3: Mean genetic distance between and within (diagonal) subspecies of *Pinus contorta*.

Subspecies	*latifolia*	*murrayana*	*contorta*	*bolanderi*
latifolia	0.0067			
murrayana	0.0078	0.0062		
contorta	0.0116	0.0112	0.0135	
bolanderi	0.0188	0.0116	0.0242	—

Table 3-4: Non neutral† loci in *Pinus contorta* and its subspecies.

	Group	Non-Neutral Loci
Haploid data set		
	All populations	211-D 250-A 250-C
	latifolia	211-D 250-A 250-C
	murrayana	211-C 250-C 250-D 429-A 623-D
	contorta	428-A
	bolanderi	243-D
Diploid data set		
	latifolia	211-D

† Determined by the Ewens-Watterson test of neutrality.

Table 3-5: Means and ranges of $\hat{M}$ (genetic similarity) and *d* (geographic distance) for pairwise comparisons all of populations and for populations within each of the three subspecies of *Pinus contorta* (see Figure 3-6 for a graphical representation of pairwise comparisons).

Group		Mean	Range
All populations	*d* (km)	1102.89	121.85 - 2883.46
	$\hat{M}$ from $\hat{\theta}$	3.66	–32.55 - 207.63
	$\hat{M}$ from G_{ST}	3.61	0.84 - 14.1
latifolia	*d* (km)	973.12	129.81 - 2621.32
	$\hat{M}$ from $\hat{\theta}$	6.92	–32.55 - 207.63
	$\hat{M}$ from G_{ST}	4.88	1.84 - 14.10
contorta	*d* (km)	738.64	121.85 - 1479.86
	$\hat{M}$ from $\hat{\theta}$	1.15	0.56 - 2.64
	$\hat{M}$ from G_{ST}	2.07	1.06 - 4.41
murrayana	*d* (km)	364.69	193.87 - 634.04
	$\hat{M}$ from $\hat{\theta}$	2.72	1.5 - 5.49
	$\hat{M}$ from G_{ST}	4.40	2.67 - 7.91

Table 3-6: Estimates of slope (*b*) from the regression of log($\hat{M}$) (genetic similarity) on log (*d*) (geographic distance) with lower (L_b) and upper (U_b) limits generated by bootstrapping.

Group		b	L_b	U_b
All populations	$\hat{M}$ from $\hat{\theta}$	–0.19479*	–0.30445	–0.06413
	$\hat{M}$ from G_{ST}	–0.08373*	–0.15882	–0.01495
latifolia	$\hat{M}$ from $\hat{\theta}$	–0.30079*	–0.45922	–0.14368
	$\hat{M}$ from G_{ST}	–0.16164*	–0.23896	–0.08859
contorta	$\hat{M}$ from $\hat{\theta}$	–0.08729	–0.36975	0.51976
	$\hat{M}$ from G_{ST}	–0.07531	–0.33304	0.47653
murrayana	$\hat{M}$ from $\hat{\theta}$	–0.33746	–1.41448	0.43722
	$\hat{M}$ from G_{ST}	–0.27182	–1.18535	0.37931

* significantly different from zero.

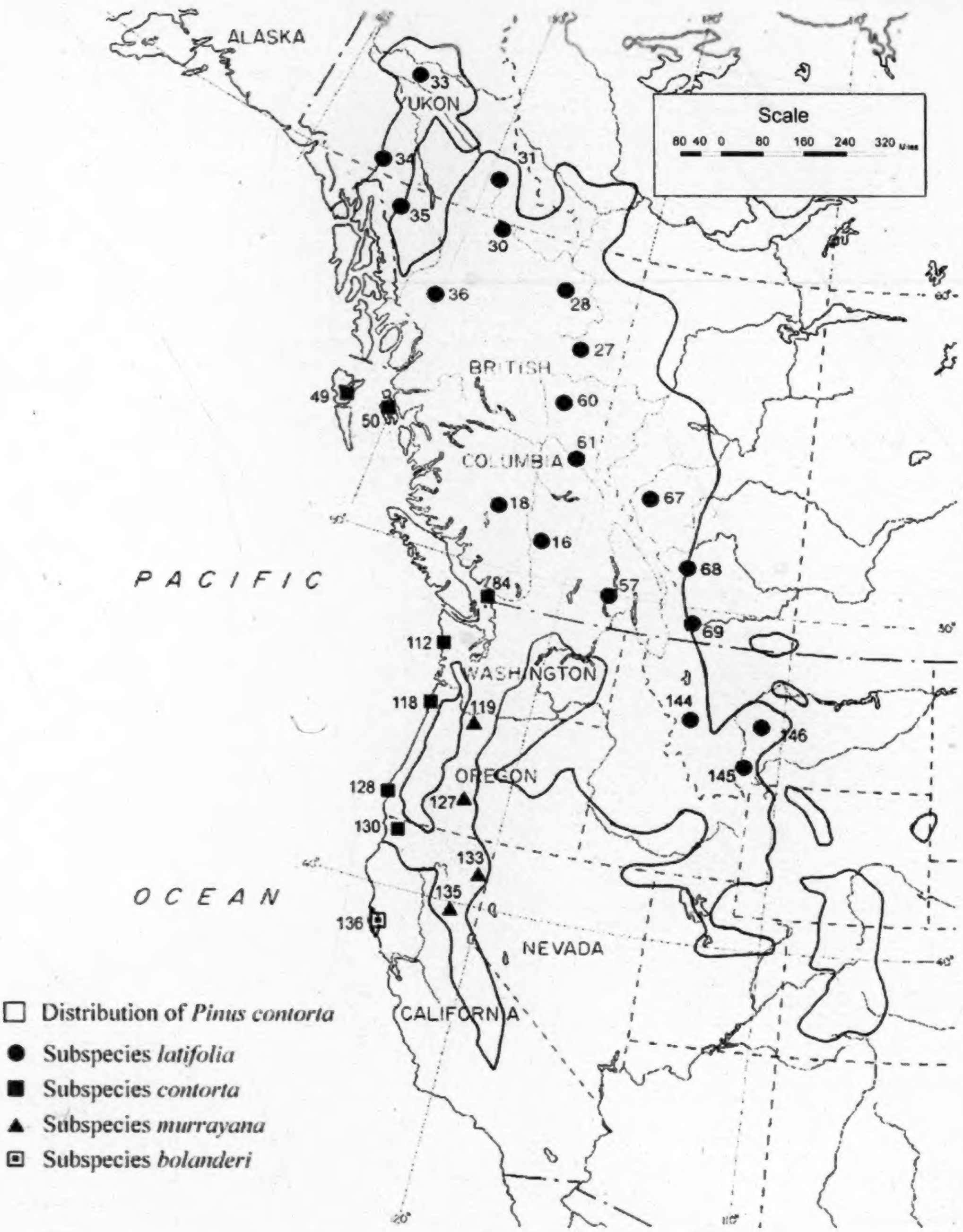
ALASKA
YUKON
BRITISH COLUMBIA
WASHINGTON
OREGON
NEVADA
CALIFORNIA
PACIFIC OCEAN
Scale
80 40 0 80 160 240 320 Miles
33
34
35
31
30
36
28
27
60
61
49
50
67
18
16
68
784
57
69
112
118
119
144
146
145
128
127
130
133
135
136
60°
50°
40°
120°
110°
☐ Distribution of *Pinus contorta*
● Subspecies *latifolia*
■ Subspecies *contorta*
▲ Subspecies *murrayana*
▣ Subspecies *bolanderi*

Figure 3-1: Location of 31 populations of *Pinus contorta* used in this study.

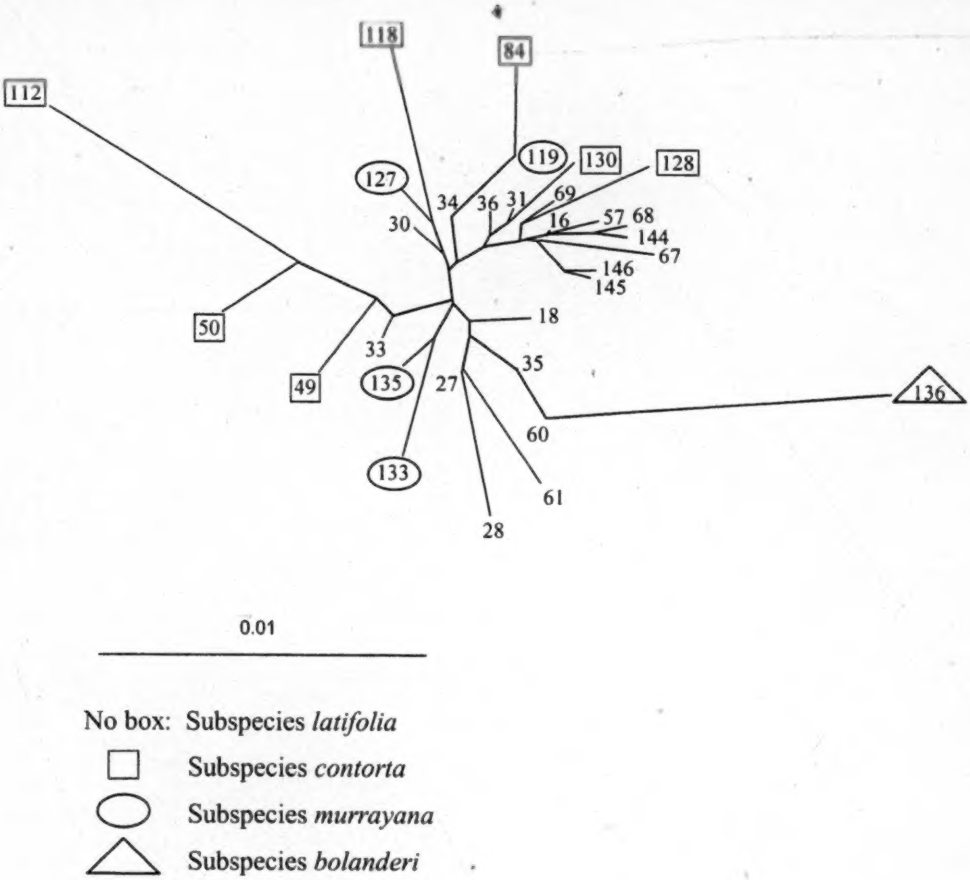
118
84
112
119
130
128
127
34
36
31
69
30
16
57
68
144
67
146
145
18
50
33
35
136
135
27
49
60
133
61
28
0.01
No box: Subspecies *latifolia*
□ Subspecies *contorta*
○ Subspecies *murrayana*
△ Subspecies *bolanderi*

Figure 3-2: Unrooted neighbour-joining tree of all 31 populations of *Pinus contorta* used in this study, based on Nei's pairwise genetic distances. Subspecies designation is after Critchfield (1957). The length of the bar indicates the genetic distance between populations.

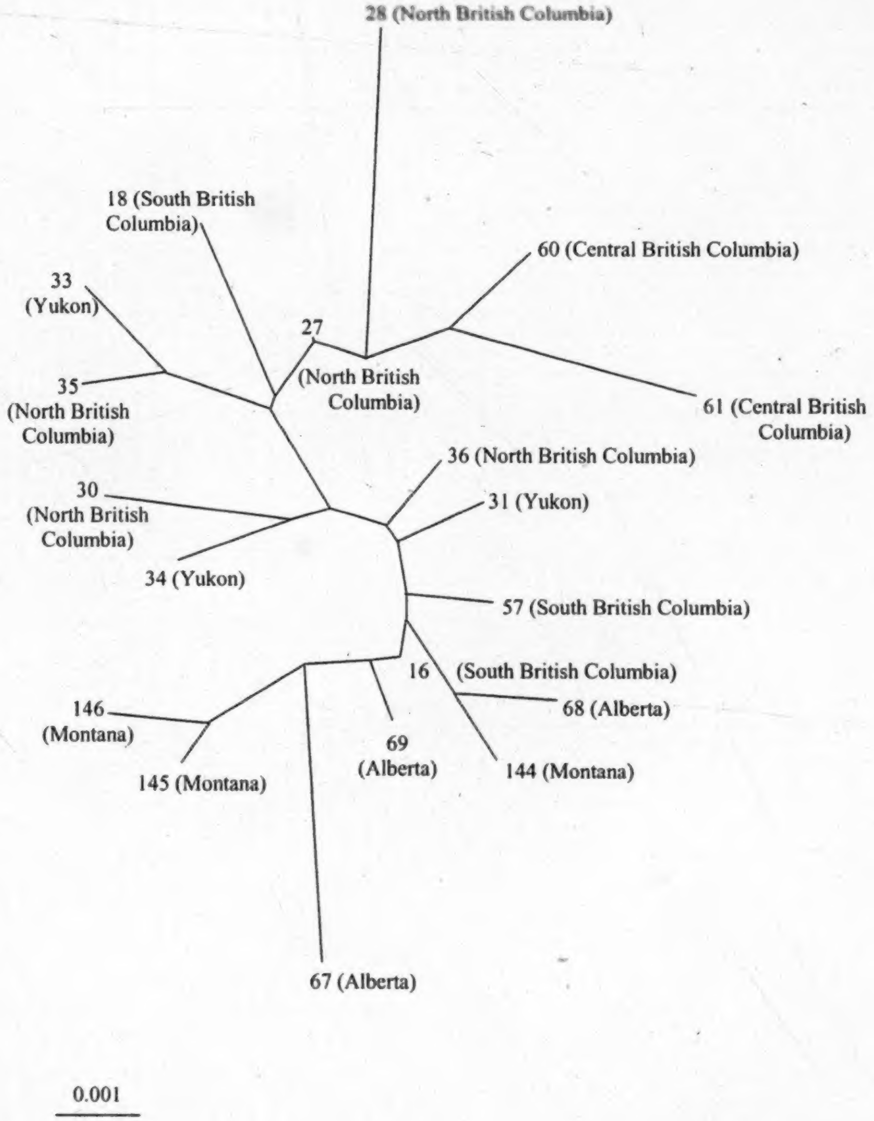
28 (North British Columbia)
18 (South British Columbia)
60 (Central British Columbia)
33 (Yukon)
27 (North British Columbia)
35 (North British Columbia)
61 (Central British Columbia)
36 (North British Columbia)
30 (North British Columbia)
31 (Yukon)
34 (Yukon)
57 (South British Columbia)
16 (South British Columbia)
146 (Montana)
68 (Alberta)
69 (Alberta)
144 (Montana)
145 (Montana)
67 (Alberta)
0.001

Figure 3-3: Unrooted neighbour-joining tree from the 19 sampled populations of *Pinus contorta* subsp. *latifolia*, based on Nei's pairwise genetic distances. Approximate geographic location is included in parentheses. The length of the bar indicates the genetic distance between populations.

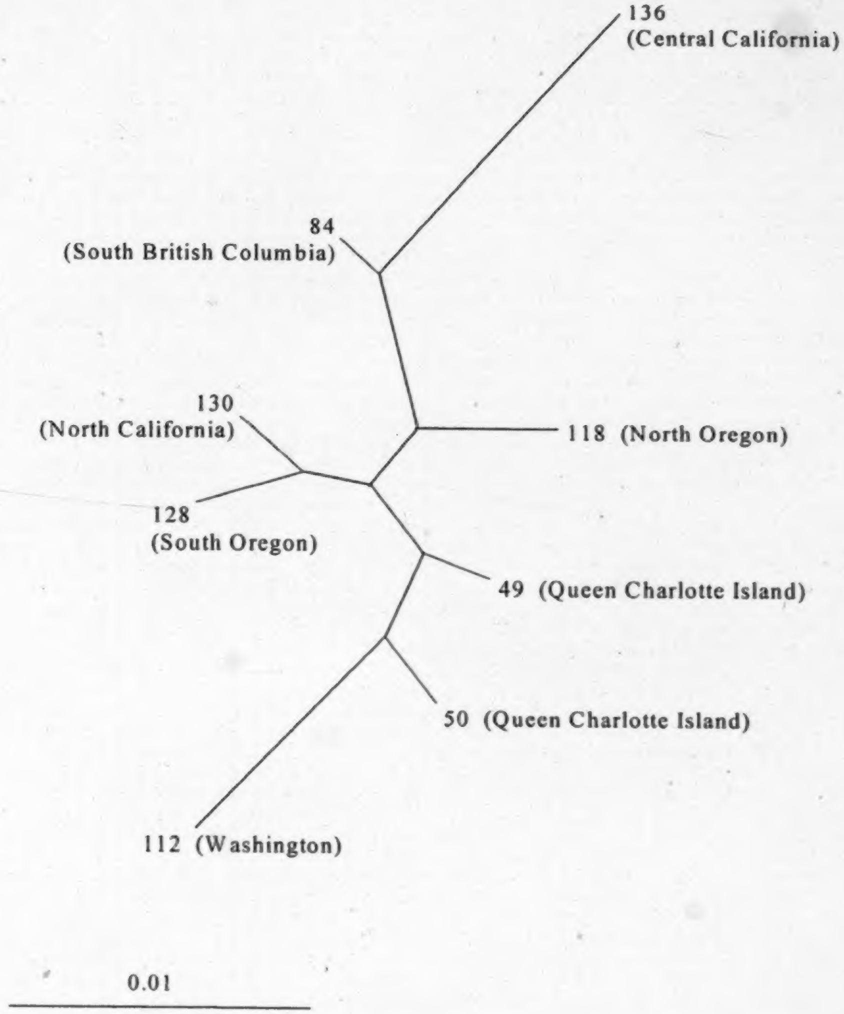
136
(Central California)
84
(South British Columbia)
130
(North California)
118 (North Oregon)
128
(South Oregon)
49 (Queen Charlotte Island)
50 (Queen Charlotte Island)
112 (Washington)
0.01

Figure 3-4: Unrooted neighbour-joining tree from the 7 sampled populations of *Pinus contorta* subsp. *contorta* and 1 sampled population of *Pinus contorta* subsp. *bolanderi* (136), based on Nei's pairwise genetic distances . Approximate geographic location is included in parentheses. The length of the bar indicates the genetic distance between populations.

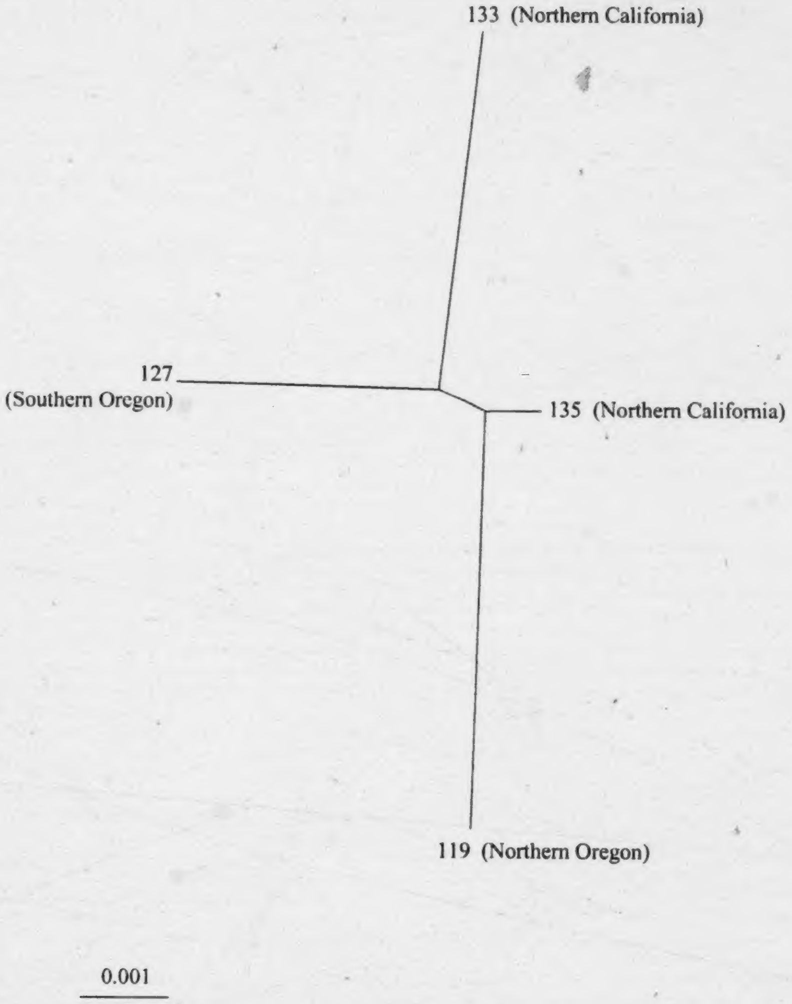
133 (Northern California)
127
(Southern Oregon)
135 (Northern California)
119 (Northern Oregon)
0.001

Figure 3-5: Unrooted neighbour-joining tree from the four sampled populations of *Pinus contorta* subsp. *murrayana*, based on Nei's pairwise genetic distances. Approximate geographic location is included in parentheses. The length of the bar indicates the genetic distance between populations.

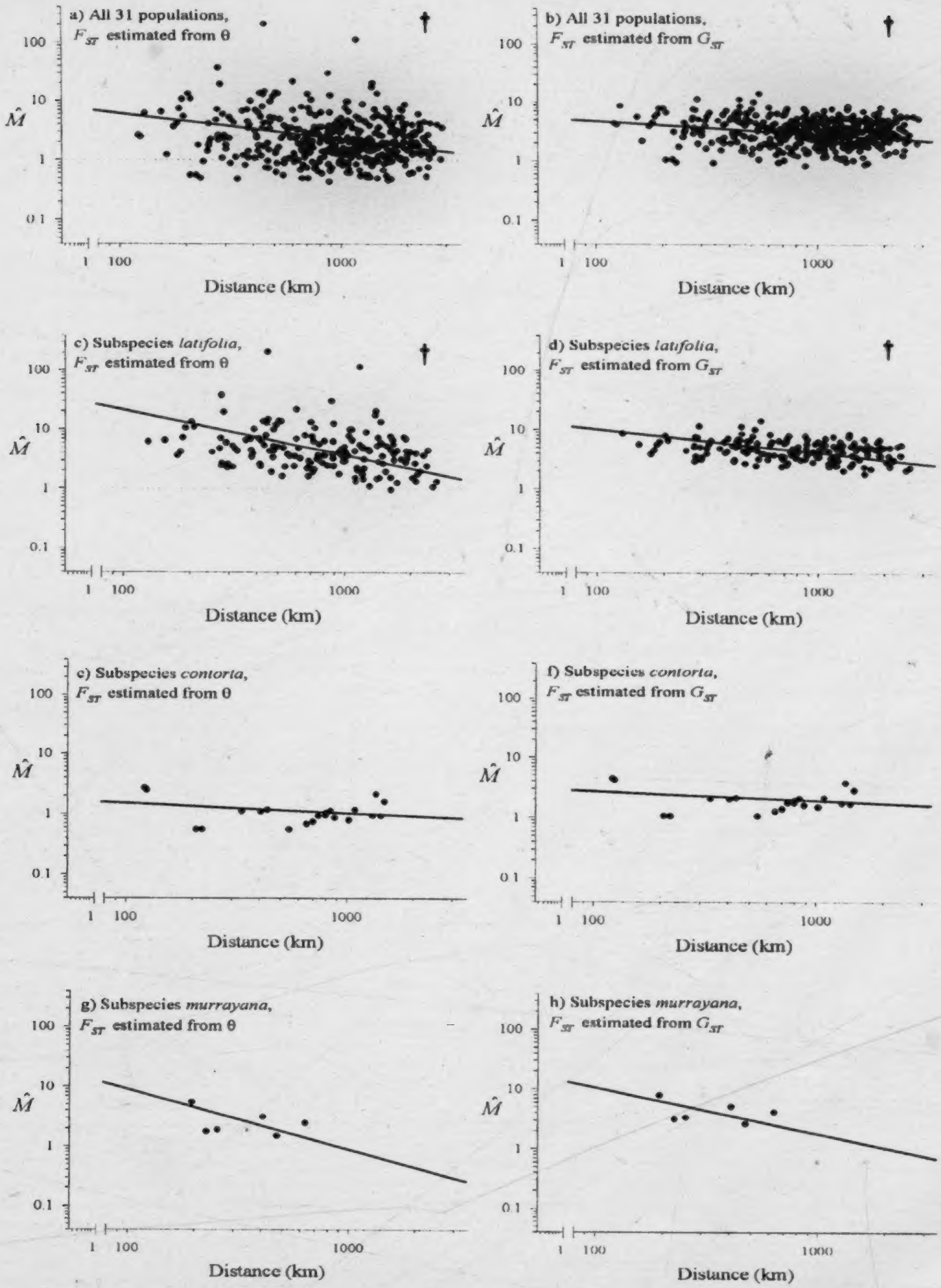

Figure 3-6: $\hat{M}$ plotted against d on a log-log scale for pairwise comparisons of all 31 populations and for populations within each of the three subspecies in *Pinus contorta*. Scatterplots are shown for $\hat{M}$ calculated using both G_{ST} and $\hat{\theta}$ as estimators of F_{ST} from the equation $\hat{M} = (1/F_{ST} - 1)/4$.

† b significantly different from zero.

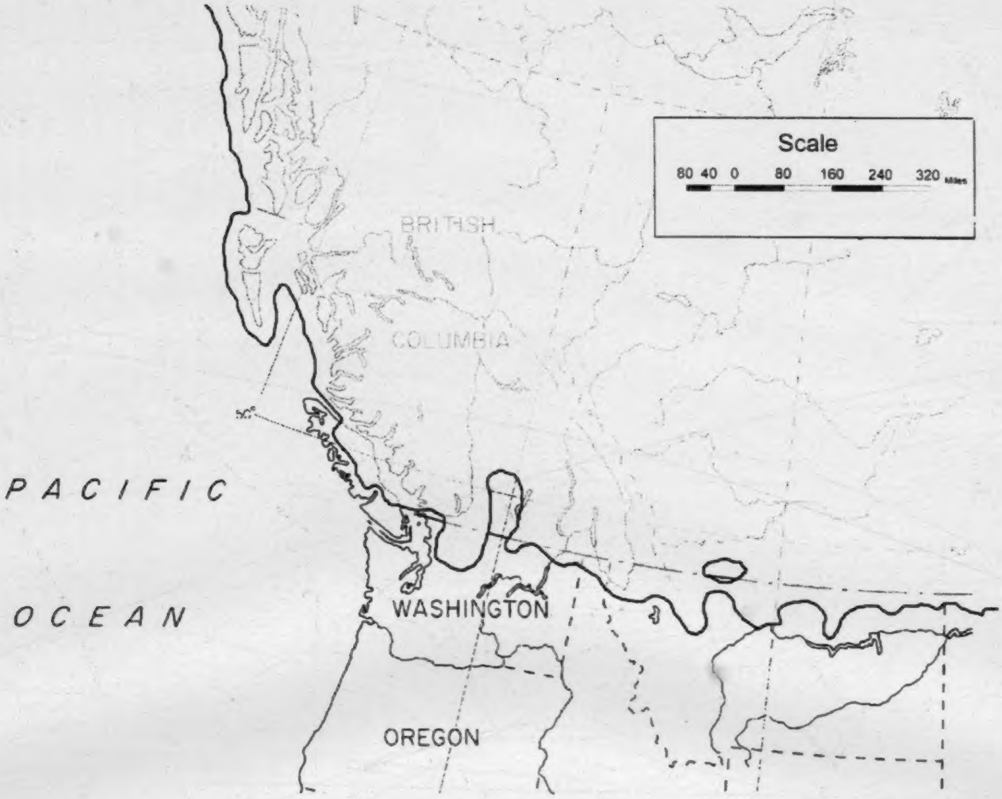
Scale
80 40 0 80 160 240 320 Miles
BRITISH
COLUMBIA
PACIFIC
OCEAN
WASHINGTON
OREGON

Figure 3-7: Approximate extent of glacial ice cover at 18,000 ybp. (After Dyke and Prest 1987).

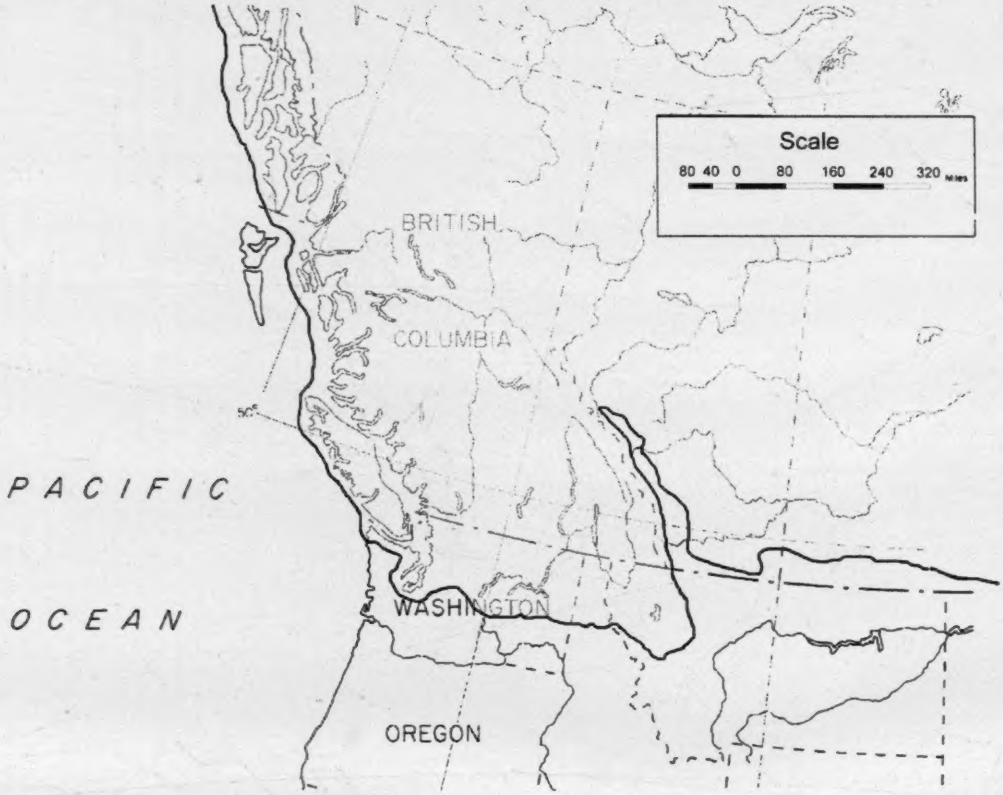
Scale
80 40 0 80 160 240 320 Miles
BRITISH
COLUMBIA
50°
PACIFIC
OCEAN
WASHINGTON
OREGON

Figure 3-8: Approximate extent of glacial ice cover at 14,000 ybp. (After Dyke and Prest 1987).

Chapter 4

Organellar sequence diversity in *Pinus contorta* Dougl.

Introduction

The study of the various modes of speciation, their causes and resulting effects, has been a cornerstone in the development of evolutionary theory. Although questions remain regarding the rate at which species divergence occurs, as well as the potential mechanisms involved, it is widely accepted (Coyne 1992) that allopatric speciation is the primary method by which species originate. Allopatric speciation occurs when populations become geographically isolated due to the introduction of a physical barrier, resulting in the evolution of reproductive barriers.

Since the initial development of reliable methods for sequencing DNA, there has been tremendous growth in the quantity of sequence information. Sequence data have been determined from across the entire spectrum of organisms, providing a wealth of evidence about species relationships and origins.

In determining relationships between closely related species, non-coding regions of DNA are utilized much more than regions that code for a specific gene. Non-coding regions are generally under reduced or no selective pressure compared with coding regions. Thus they are able to accumulate substitutions at individual nucleotide sites, as well as insertions and deletions, faster than coding regions, because many mutations in coding regions tend to be eliminated by selection.

The *trnL/F* region of the chloroplast genome has been used for a large number of studies and has been shown to be very useful in determining phylogeny among related species (e.g. Gielly *et al.* 1996; Bellstedt *et al.* 2001; Nyffeler 2002; Nishikawa *et al.* 2002; Gonzalez and Vovides 2002). When transcribed, the *trnL* (UAA) and *trnF* (GAA) genes produce two transfer RNAs, which are utilized for transferring amino acids (leucine and phenylalanine, respectively) to the ribosome during protein synthesis. These genes are separated by a short spacer segment of DNA, approximately 400 base pairs in length, depending on species. The *trnL* gene also contains an intron of approximately 500 bp in length. Due to the functional importance of transfer RNA in protein synthesis, these genes are highly conserved across species. As a result, primers designed by Taberlet *et al.* (1991), which are anchored in the genes, are able to amplify this region across a wide variety of species.

Demesure *et al.* (1995) developed additional universal primers for non-coding regions of mitochondrial and chloroplast DNA. Starting with two of these primers, Latta and Mitton (1999) showed that the b/c intron of the mitochondrial *nad1* gene had size variation in amplification fragments of *Pinus ponderosa* Laws (ponderosa pine). This intron contains an area of two repeated elements that are 34 and 32 base pairs in length. Variation in the number of these elements has been observed in *Pinus ponderosa*, *Pinus flexilis* James (limber pine) and *Pinus sylvestris* L. (scots pine) (Latta and Mitton 1999; Mitton *et al.* 2000a; Mitton *et al.* 2000b).

Lodgepole pine (*Pinus contorta* Dougl.) is an important component of forest ecosystems in western North America. Its current range extends from the Yukon, through British Columbia and western Alberta, south to California and southeast along the Rocky Mountains into Idaho, Utah, and Colorado. This large distribution over extremely varied habitats can be attributed to its tolerance of a wide range of environmental conditions. It has been shown (Dong *et al.* 1992; Dong and Wagner 1994) that *Pinus contorta*, like other conifers (Brent and David 1989, Neal and Sederoff 1989), has a paternally inherited chloroplast genome. Mitochondria however, are maternally inherited although paternal leakage of mitochondria has been observed at low frequency in *Pinus* (Wagner *et al.* 1991). Thus, in a wind-pollinated outcrossing species such as lodgepole pine, differences are expected in the rate of gene flow between the chloroplast and mitochondrial genomes. The lighter pollen has a much greater potential for dispersal and corresponding potential for a greater level of gene flow compared to the mitochondrial genome. Differential rates of gene flow between the two organellar genomes have been detected in other pine species (Dong and Wagner 1994; Latta and Mitton 1997; Latta and Mitton 1999).

Critchfield (1957) subdivided *Pinus contorta* into four subspecies. *Pinus contorta* subsp. *contorta* is a coastal variety (Figure 4-1), restricted to areas proximal to the Pacific Ocean. *Pinus contorta* subsp. *murrayana* ranges from the Cascade Mountains of Oregon south to the Sierra Nevada. *Pinus contorta* subsp. *latifolia* is the most widespread of the subspecies, occupying the Rocky Mountain range and interior British Columbia. *Pinus contorta* subsp. *bolanderi* is restricted to a very small area on the coast of California near Mendocino Bay.

The majority of the current range of *Pinus contorta* was covered by ice during the last glaciation period, the Wisconsin (Dyke and Prest 1987). As the ice retreated, opportunities for colonization of newly exposed land existed for species in adjacent unglaciated areas. The morphological differences between the subspecies of *Pinus* *contorta* have led to the established belief that they have been separated genetically for a considerable time (Wheeler and Guries 1982; Critchfield 1985), and geographically isolated during the Pleistocene (Critchfield 1985). If this is the case, then it is probable that the distribution of sequence variation should reflect this, supporting the division of lodgepole pine into its four subspecies.

The objective in this study was to utilize sequence data from both chloroplast and mitochondrial DNA in *Pinus contorta* to attempt to determine if the geographic range of the various subspecies corresponds with patterns of genetic variation and to compare the relative amount of information generated from each of the two genomic regions.

Materials and methods

Population sampling

Needle tissue was collected from provenance plantations established by the British Columbia Ministry of Forests at Prince George and at Lake Cowichan. The provenances at these locations represent populations from much of the range of *Pinus contorta* (Figure 4-1). Representatives of each population were sampled to a maximum of five individuals (Table 4-1). The individual trees sampled are numbered according to the population number and position within the provenance trial. Some provenances have less than five individuals remaining due to low survival rates. Samples were placed in a cooler with dry ice during collecting and later stored at –20° C. A total of 31 populations were represented by 140 individuals. Population 136 (subsp. *bolanderi*) had not survived at either of the provenance plantations. Subsequent to the needle collections an attempt was made to germinate seeds from this population that had been collected previously, and stored at –20° C. Germinated seedlings were grown for approximately six weeks before being harvested for DNA extraction.

DNA extraction

DNA extraction was similar to the methods of Doyle and Doyle (1990). One gram of needle tissue was briefly rinsed with water and then ground to a fine powder in liquid nitrogen. The ground sample was transferred to an Oakridge tube, which was then filled with extraction buffer (50 mM Tris pH 8, 5 mM EDTA, 0.35 M Sorbitol). The tube was then centrifuged at 9,000 rpm for 15 minutes at 4° C. The supernatant was removed and the pellet re-suspended in 5 ml of extraction buffer. One ml of 5% N-lauryl sarcosine was added and mixed, after which the sample was incubated for 30 minutes at room temperature. One ml of 5M NaCl, and 0.8 ml of CTAB solution (10% CTAB, 0.7 M NaCl) was added and the sample was then incubated for 10-15 minutes in a 60° C water bath. An equal volume (approximately 10 ml) of chloroform/isoamyl alcohol was added and gently mixed. The sample was then centrifuged at 10 krpm for 10 minutes. A transferpette was used to transfer the supernatant to a new Oakridge tube. Twenty ml of 100% ethanol and 200 µl of 3M sodium acetate were added and the sample stored at –20° C overnight. The sample was then centrifuged at 6,000 rpm for 20 min at 4° C to pellet the DNA. The pellet was transferred to a 1.5 ml microcentrifuge tube and washed twice with 76% ethanol/10 mM ammonium acetate. After allowing the pellet to dry, it was re-suspended in TE buffer. 5 to 10 µl of RNase A (stock solution 0.1 µg/µl) was added and the sample incubated at 36° C for 1 hour after which the sample was stored at –20° C.

Amplification and sequencing

Two adjacent chloroplast regions, (the *trnL* intron, and the *trnL/trnF* spacer) and one mitochondrial region (the *nad1* b/c intron) were selected for sequencing, using the primers of Taberlet *et al.* (1991) and Mitton *et al.* (2000b).

Chloroplast region:

Primers c and f (Table 4-2) were used to amplify a fragment approximately 1,100 base pairs in length from individuals of *Pinus contorta*. The amplification reaction conditions were as follows: a 25 µl volume containing 10mM Tris-HCl pH 8.3, 50mM KCl, 2.5 mM $MgCl_2$, 200µM of each of the four dNTPs, 0.5µM of each primer, 0.5 units *Taq* polymerase (Perkin-Elmer), and 20-40ng DNA was subjected to 35 cycles of 1 min. 95° C, 1 min. 55° C, 2 min. 72° C, after an initial denaturation of step of 2 min. at 95° C. After the 35 cycles an elongation step of 5 min. at 72° C was performed, followed by a 4° C soak. Amplification of DNA was carried out in a GeneAmp 9600 Thermo-cycler (Perkin-Elmer). Excess primers and unincorporated nucleotides in the amplification tube were removed using PCR Product Pre-Sequencing kit (USB Corporation).

A simultaneous bi-directional sequencing strategy was employed on a Li-Cor 4200 sequencer. The amplified fragment was sequenced directly using fluorescently labelled primers. Two sets of primers were used for sequencing, one set for each of the regions of interest (Table 4-2). Sequencing reactions were done using the SequiTherm EXCEL II DNA sequencing kit (Epicentre) following manufacturer's instructions, and loaded onto the Li-Cor gel rig. By labelling both the forward and reverse sequencing primers with different fluorescent dyes, the fragment was sequenced in both directions in the same reaction. Due to unsatisfactory results from primer f, a second reverse amplification primer (trnF-Rb) was designed, located 10 base pairs downstream of primer f. By nesting sequencing primer f in this fashion, improved sequence data were obtained.

Mitochondrial region:

The primers published by Mitton *et al.* (2000b) (primers NAD1B1F and NAD1C1R) were used to successfully amplify a fragment approximately 1,537 base pairs in *Pinus contorta*. Primers NAD1B3F and NAD1C3R were used to directly sequence a 350 base pair region of the fragment for all samples of *Pinus contorta* utilizing the simultaneous bi-directional sequencing strategy in a similar fashion to the two chloroplast sequences.

Bases were identified using software provided by Li-Cor and chromatograms were generated for each sample. Chromas 2.13 (Technelysium) was used to confirm the data obtained from the forward and reverse primers.

Data analysis

Sequences were aligned using ClustalX 1.8 (Thompson *et al.* 1997). The two chloroplast data sets were pooled together for analysis, since they show a similar degree of differentiation among populations, and are physically close (genetically completely linked). Measures of genetic distance were calculated using the method of Tajima and Nei (1984) within and between subspecies using MEGA version 2.1 (Kumar *et al.* 2001). Estimates of genetic distance using other methods yielded almost identical results. DnaSp version 3.53 was used to estimate nucleotide diversity from Nei (1987), and to perform tests of neutrality. The algorithms used by these programs estimate nucleotide diversity. Variation in length of the region due to the presence of insertions or deletions (indels) will therefore not contribute to these estimates. Maximum parsimony analyses were conducted on the combined chloroplast data sets using the close-neighbour interchange option in MEGA with uniform weighting. For these analyses, sequence data of the chloroplast *trnL* intron and *trnL/F* spacer regions from *P. thunbergii* (GenBank Accession No. NC 001631) was used as an outgroup.

Results

The *trnL* intron of *Pinus contorta* is 487 base pairs long. At this locus, five polymorphic nucleotide sites were detected. Additionally, four individuals, two from population 49, and one each from populations 95 and 36 shared a short indel of 5 bp.

The *trnL/trnF* spacer region had four polymorphic sites among 385 sites. A single bp indel was observed in one individual from population 135. A 26 bp deletion was also observed in one individual from population 31. For both chloroplast regions, the polymorphic nucleotide sites were confined to a very small number of individuals (Table 4-3, Table 4-4). The great majority of individuals displayed a uniform sequence for each of the regions. This sequence homogeneity is reflected in the overall mean diversity of 0.000178 for the intron and 0.000186 for the spacer. The degree of differentiation of the subspecies based on genetic distance (Table 4-5) is also very small.

Parsimony analyses of the data yielded 8,610 equally parsimonious trees (Figure 4-3). The strict consensus tree is presented in Figure 4-4. Due to the large amount of uniformity in the sequence data, individuals with uniform sequences are represented together in these figures by the category 'All others'.

Two tests of neutrality were performed on the chloroplast data. Both Tajima's test (Tajima 1989) and Fu and Li's test (Fu and Li 1993) were significant (D = –2.26, and –4.52 respectively), indicating a departure from neutral evolution.

The two repeated elements found in *Pinus ponderosa* are also present in *Pinus contorta*, although each element differs at one nucleotide. The region of repeated DNA from *Pinus contorta* is compared with *Pinus ponderosa* in Figure 4-2. The number of repeated elements is fewer in *Pinus contorta* and the last element is interrupted by an indel. Unlike *Pinus ponderosa*, which displays geographic variation in the number of repeated elements, no differences in the number of repeats were observed between any of the samples of *Pinus contorta*.

Discussion

The most obvious result of the study is the lack of differentiation detected among individuals of *Pinus contorta* for the chloroplast DNA and mitochondrial DNA regions surveyed. Although it is expected that wide distributions of haplotypes would be observed in a highly outcrossing, wind-pollinated species such as *Pinus contorta*, it was unexpected that only one haplotype would be widely distributed.

The *trnL*/*trnF* region has been shown to be useful for phylogeny reconstruction at the species level (e.g. Gielly *et al.* 1996; Nishikawa *et al.* 2002; Gonzalez and Vovides 2002) and determining hybrid parentage (Chen *et al.* 2002). At the subspecies level in *Pinus contorta* however, the level of resolution is insufficient to separate the four subspecies.

The lack of differentiation in the mitochondrial *nad1* b/c intron sequence data of *Pinus contorta* is in sharp contrast with data obtained from *Pinus ponderosa* and *Pinus flexilis*. For both of these species, differences in repeat number in the indel region were useful for revealing geographic structure. *Pinus ponderosa* and *Pinus contorta* are thought to have shared a similar biogeographical history. *Pinus ponderosa* existed in two large groups south of the ice-sheet during the Wisconsin (~ 100,000 years duration) that are thought to have been geographically isolated from one another (Conkle and Critchfield 1988). During the same time period, *Pinus contorta* was widespread south of the ice sheet (Critchfield 1985). The morphological distinctness of each of the races of *Pinus contorta* led Critchfield (1985) to suggest that these four groups have been genetically isolated for millennia. Latta and Mitton (1999) showed that data from chloroplast and mitochondrial sequences each displayed three haplotypes that correspond strongly with the historical east-west division in *Pinus ponderosa*. The homogeneity of the data presented here from *Pinus contorta* indicates that either the history of *Pinus contorta* is somewhat different that previously thought, or that the particular loci examined are not informative in *Pinus contorta*.

Since the *nad1* b/c intron is monomorphic in *Pinus contorta*, the remainder of the discussion will focus on the data obtained from the two chloroplast regions.

The most striking result from the chloroplast data was the widespread distribution of the most common haplotype, which was present in all the sampled populations. For most populations this was the only haplotype present. Single nucleotide polymorphisms seemed to be distributed somewhat randomly, although it can be seen that populations with such individuals are all located towards the margins of the species range. Since marginal populations are often smaller in number this result is the opposite of what is expected. The expected number of alleles increases with population size under a constant rate of mutation (Hartl and Clark 1989). The distribution of variation in this manner suggests that other factors must be involved. In the northern part of the range, rare alleles have been detected in allozymes (Yeh and Layton 1979; Wheeler and Guries 1982) and in RAPDs (Chapter 3). The presence of rare alleles in this area has been attributed to persistence of *Pinus contorta* in a Yukon refugium during the Wisconsin (Wheeler and Guries 1982), and the pooling of genes from introgression of subsp. *contorta* and *latifolia* (Chapter 3). Rare alleles in the eastern and southern portions of the species range may be a result of secondary contact with *Pinus banksiana*. Additionally, these populations have likely existed for a greater amount of time than populations in the central portions of the range, and thus have had more opportunity to accumulate mutations.

A comparison of the chloroplast data between *Pinus contorta* and *Pinus ponderosa* suggests two possibilities. First, the loci chosen in this study may simply not be reflective of an historical separation of subspecies in *Pinus contorta*. Plant organellar genomes exhibit lower synonymous nucleotide substitution rates than plant nuclear or animal genomes (Li 1997). Although plant mitochondrial DNA does undergo frequent re-arrangements, including large scale duplications and deletions, sampling over a number of loci may be required to detect variation.

Alternately, the differences in the distribution of variation may be a result of some differences in the histories of the two species. The division of *Pinus ponderosa* into two geographic groups during the course of the Wisconsin allowed different chloroplast (and mitochondrial) haplotypes to become widespread in each location (Latta and Mitton 1999). The lack of differentiation observed between subspecies of lodgepole pine may indicate that such a division may did not exist, or was more poorly defined in *Pinus contorta*.

Given that previous glaciations of this ice age were of similar extent and duration as the Wisconsin, it is reasonable to predict these would have a similar effect on the distribution of species. The divergence seen between the subspecies of *Pinus ponderosa* may be a result of repeated separation with only limited contact during interglacial periods. The current degree of introgression between the subspecies is localized to a small geographic area (Latta and Mitton 1999). If the cycle continues, glacial advance during the next 1,000 to 2,000 years (lasting for ~100,000 years) would again cause the two subspecies to be separated, permitting continued divergence. Repeated glaciation may have had the opposite effect on *Pinus contorta*. A reduction of the distribution to areas south of the ice during glacial advance would have reduced the size of subsp. *latifolia* greatly. A reduction of population size could have resulted in reduced variation due to the effects of genetic drift.

The analysis of the variance in G_{ST} from Chapter 3 indicates that the allele frequencies of the subspecies have not drifted apart. Much more heterogeneity in values of G_{ST} is seen within each of the subspecies than is seen between them. This is what would be expected if the separation of the subspecies were too recent to be detected with RAPD markers.

If a lack of divergence between the subspecies in *Pinus contorta* is a result of significant levels of gene flow that existed between populations that survived south of the glacial terminus, why are the subspecies so morphologically diverse? Large differences in morphology do not necessarily imply a large genetic distance at neutral loci between populations or even between species. A good example of discordance between these two types of data is the Hawaiian silversword alliance (Witter and Carr 1988; Baldwin *et al.* 1990). The morphological differences seen in the subspecies of *Pinus contorta* today may have developed very quickly in response to changing climate and associated selection pressure. Cone serotiny for example is the primary trait distinguishing subsp. *murrayana* and *latifolia*. As climate changed during the course of glacial retreat, the frequency and pattern of fire events may have also changed, influencing the degree of cone serotiny in particular populations. Such selection pressure combined with the founder effects associated with the rapid expansion of populations may be the primary source of morphological differences between these subspecies observed today. The demarcation of subsp. *murrayana* and *latifolia* is, in fact, not very clear. These two subspecies show a continuous cline in a number of characters (Critchfield 1957). A population designated as subsp. *murrayana* (based on morphology) near Klamath Falls in southern Oregon was much closer in terms of genetic distance to populations of subsp. *latifolia* in northern Washington and south-eastern British Columbia than to other populations of subsp. *murrayana* (Wheeler and Guries 1982). Interestingly, Dong and Wagner (1993) also reported a close affinity between a central Oregon population and a central British Columbia population of *Pinus contorta*.

In an allozyme study of the relationship between subsp. *contorta* and *bolanderi*, Aitken and Libby (1994) found that the average F_{ST} between subspecies was only marginally greater than F_{ST} within populations of each taxon. This contrasts with the large differences in morphological, survival, and growth traits observed between the two subspecies. Aitken and Libby (1994) provide a range for the time of divergence between the two subspecies of 18,000 years ago (from allozymes) to 3,000 – 8,000 years ago [from the palynological data of Huesser (1960)]. The remarkable divergence in morphology over this short time period without a corresponding divergence in genetic markers indicates the potential for similar occurrences between the other subspecies.

The presence of an indel in the *trnL* intron localized to populations on the Queen Charlotte Islands and in two nearby populations is interesting. Failure to detect this mutation in populations south of this area indicates that it has likely occurred subsequent to colonization of the Queen Charlottes by subsp. *contorta* after glacial retreat. Populations 95 and 36 also share this indel. Given that the location of these two populations were covered in ice for several thousand years while the Queen Charlotte Islands were ice-free (Dyke and Prest 1987), and that the prevailing wind patterns in this area are north and east, migration or gene flow from the Queen Charlottes to coastal British Columbia to interior British Columbia is suggested.

The fact that the 5 bp indel in the *trnL* intron was shared among individuals of both subsp. *contorta* and *latifolia* provides some evidence of common ancestry or gene flow between the two subspecies and supports the hypothesis (Chapter 3) that the distinctness observed in Yukon and north British Columbia populations (Yeh and Layton

1979; Wheeler and Critchfield 1982) from other populations of subsp. *latifolia* is due to migration of genes from north coastal populations into the interior of British Columbia.

In the Yukon, at the north end of the range of *Pinus contorta*, it has been postulated that ice-free areas supported small populations of this species throughout the Pleistocene. Pollen data from deposits in this area indicate the presence of *Pinus contorta* in pre- or early Pleistocene to the north and west of its current distribution (Matthews 1970, Lichi-Federovich 1974), however the absence of pine is indicated in Alaska throughout the Holocene (Hopkins *et al.* 1981). Hopkins (1979) did not believe that the climatic conditions during the Pleistocene in this area could support pine. From a series of sediment cores taken from southwestern Alberta to the Yukon, MacDonald and Cwynar (1985) showed a steady migration of *Pinus contorta* from the south, arriving near the current limits of its distribution in central Yukon 400 years ago. The lack of divergence between populations in the Yukon and populations to the south, at the loci studied here, is likely a reflection of the extremely slow rate of change of plant organellar genomes rather than an indication that a Yukon refugium did not exist.

Both of the tests of neutrality performed were significant at $p < 0.02$. The significance of D in Tajima's test indicates a condition of purifying selection or recent population expansion. Although the regions sequenced here are from non-coding DNA, they are completely linked to coding regions in the organelle. Thus, either or both conditions may explain the significant value for D. Fu and Li's test is based on a comparison of number of mutations in external versus internal branches of a phylogenetic tree. Although the number of branches is small in this data set, the two tests have similar results.

Conclusion

Although the loci investigated here have proved useful for phylogeny reconstruction and studies of biogeography in other species they show little differentiation in *Pinus contorta*. This may be a result of the slow rate of evolution of the loci studied here. Alternately, the small amount of variation observed may be a result of fixation of alleles due to reduced population size during the Wisconsin and the homogenizing effect of gene flow. Expansion of populations after glacial retreat would have resulted in the proliferation of the most common alleles, particularly if the populations advanced in the stepwise manner indicated by MacDonald and Cwynar (1985) and Cwynar and Macdonald (1987). Distinguishing between these alternatives would require additional sampling at other loci. Although the mode of inheritance for the chloroplast and mitochondrial genomes is different in *Pinus contorta*, comparisons of the relative amount of gene flow from pollen versus seed dispersal, using data from these genomes, will be confounded with the difference in the rate of change of each genome.

References

Aitken, S.N. and W.J. Libby. 1994. Evolution of the pygmy-forest edaphic subspecies of *Pinus contorta* across and ecological staircase. Evolution **48**: 1009-1019.

Baldwin, B.G., D.W. Kyhos and J. Dvorák. 1990. Chloroplast DNA evolution and adaptive radiation in the Hawaiian silversword alliance (Asteraceae-Madiinae). Ann. Missouri Bot. Gard. **77**: 96-109.

Bellstedt, D.U., H.P. Linder and E.H. Harley. 2001. Phylogenetic relationships in *Disa* based on non-coding *trnL-trnF* chloroplast sequences: evidence of numerous repeat regions. Am. J. Bot. 88: 2088-2100.

Brent, H.M. and D.E. David. 1989. Unusual inheritance patterns of organelle DNA in conifers. Tibtech. **7**: 257-258.

Chen, J., C.G. Tauer and Y. Huang. 2002. Paternal chloroplast inheritance patterns in pine hybrids detected with *trnL–trnF* intergenic region polymorphism. Theor. Appl. Genet. **104**: 1307-1311.

Chromas 2.13. Technelysium Pty Ltd. Helensvale Queensland Australia.

Coyne, J.A. 1992. Genetics and speciation. Nature **355**: 511-515.

Critchfield, W.B. 1957. Geographic variation in *Pinus contorta*. Maria Moors Cabot Foundation. Publ. 3. Harvard University Press.

Critchfield, W.B. 1985. The late Quaternary history of lodgepole and jack pines. Can J. For. Res. **15**: 749-772.

Cwynar, L.C. and G.M. MacDonald. 1987. Geographical variation of lodgepole pine in relation to population history. Am. Nat. **129**: 463-469.

Dong, J., D.B. Wagner, A.D. Yanchuk, M.R. Carlson, S. Magnussen, X. Wang and A.E. Szmidt. 1992. Paternal chloroplast DNA inheritance in *Pinus contorta* and *Pinus banksiana*: independence of parental species or cross direction. J. Hered. **83**: 419-422.

Dong, J. and D.B. Wagner. 1994. Paternally inherited chloroplast polymorphism in *Pinus*: estimation of diversity and population subdivision and tests of disequilibrium with a maternally inherited mitochondrial polymorphism. Genetics **136**: 1187-1194.

Doyle, J.J. and J.L. Doyle. 1990. Isolation of plant DNA from fresh tissue. BRL Focus **12**: 13-15.

Dyke, A.S. and V.K. Prest. 1987. Late Wisconsinan history of the Laurentide ice sheet. Geog. Phys. Quater. **41**: 327-364.

Fu, Y.-X. and W.-H. Li. 1993. Statistical tests of neutrality of mutations. Genetics **133**: 693-709.

Gielly, L. and P. Taberlet. 1996. A phylogeny of the European gentians inferred from chloroplast *trnL* (UAA) intron sequences. Bot. J. Linn. Soc. **120**: 57-75.

González, D. and A.P. Vovides. 2002. Low intralineage divergence in *Ceratozamia* (Zamiaceae) detected with nuclear ribosomal DNA ITS and chloroplast DNA *trnL-F* non-coding region. Syst. Bot. **27**: 654-661.

Heusser, C.J. 1960. Late Pleistocene environments of north Pacific North America. Am. Geogr. Soc., Special Publ. No. 35.

Hopkins, D.M. 1979. Landscape and climate of Beringia during late Pleistocene and Holocene time. *In*: The first Americans: origins affinities, and adaptations. Edited by W.S. Laughlin and A.B. Harper. Gustav Fischer, New York. pp. 15-41.

Hopkins, D.M., P.A. Smith and J.V. Matthews Jr. 1981. Dated wood from Alaska and the Yukon: implications for forest refugia in Beringia. Quat. Res. **15**: 217-249.

Kumar, S., K. Tamura, I.B. Jakobsen and M. Nei. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software, Arizona State University, Tempe, Arizona, USA.

Latta, R.G and J.B. Mitton. 1997. A comparison of population structure across four classes of gene marker in limber pine. Genetics **146**: 1153-1163.

Latta, R.G and J.B. Mitton. 1999. Historical separation and present gene flow through a zone of secondary contact in Ponderosa pine. Evolution **53**: 769-776.

Lichti-Federovich, S. 1974. Palynology of two sections of late Quaternary sediments from the Porcupine River, Yukon Territory. Can. Geol. Surv. Pap. 74-23.

MacDonald, G.M. and L.C. Cwynar. 1985. A fossil pollen based reconstruction of the late Quaternary history of lodgepole pine (*Pinus contorta* ssp. *latifolia*) in the western interior of Canada. Can J. For. Res. **6**: 1039-1044.

Matthews, J.V. Jr. 1970. Quaternary environmental history of interior Alaska: pollen samples from organic colluvium and peats. Arct. and Alp. Res. **2**: 241-251.

Mitton, J.B., B.R. Kreiser and R.G. Latta. 2000a. Glacial refugia of limber pine (*Pinus flexilis* James) inferred from the population structure of mitochondrial DNA. Mol. Ecol. **9**: 91-97.

Mitton, J.B., B.R. Kreiser and G.E. Rehfeldt. 2000b. Primers designed to amplify a mitochondrial *nad1* intron in ponderosa pine, *Pinus ponderosa*, limber pine, *P. flexilis*, and Scots pine, *P. sylvestris*. Theor. Appl. Genet. **101**: 1269-1272.

Neale, D.B. and R.R. Sederoff. 1989. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. Theo. Appl. Genet. **77**: 212-216.

Nyffeler, R. 2002. Phylogenetic relationships in the cactus family (Cactaceae) based on evidence from *trnK/ matK* and *trnL-trnF* sequences. Am. J. Bot. **89**: 312-326.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia Univ. Press, New York.

Nishikawa, T., B. Salomon, T. Komatsuda, R. von Bothmer and K.-i. Kadowaki. 2002. Molecular phylogeny of the genus *Hordeum* using three chloroplast DNA sequences. Genome **45**: 1157-1166.

Peteet, D.M. 1991. Postglacial migration history of lodgepole pine near Yakutat, Alaska. Can. J. Bot. **69**: 786-796.

Rozas, J. and R. Rozas. 1999. DnaSp version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15**: 174-175.

Taberlet, P., L. Gielly, G. Pantou and J. Bouvet. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Bio. **17**: 1105-1109.

Tajima, F. and M. Nei. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. **1**: 269-285.

Wagner, D.B., J. Dong, M.R. Carlson and A.D. Yanchuk. 1991. Paternal leakage of mitochondrial DNA in *Pinus*. Theor. Appl. Genet. **82**: 510-514.

Wheeler, N.C. and R.P. Guries. 1982. Biogeography of lodgepole pine. Can J. Bot. **60**: 1805-1814.

Whitter, M.S. and G.D. Carr. 1988. Adaptive radiation and genetic differentiation in the Hawaiian silversword alliance (compositae: madiinae). Evolution **42**: 1278-1287.

Yeh, F.C. and C. Layton. 1979. The organization of genetic variability in central and marginal populations of lodgepole pine *Pinus contorta* ssp. *latifolia*. Can. J. Genet. Cytol. **21**: 487-503.

Table 4-1: Number of trees sampled per population for 31 populations of *Pinus contorta*.

Subspecies	Population	Number of trees sampled
latifolia	26	5
	31	5
	33	4
	35	5
	36	5
	38	5
	57	5
	67	5
	69	4
	100	5
	111	5
	117	5
	141	4
	143	5
	145	5
	154	4
contorta	49	4
	50	4
	78	4
	82	4
	88	4
	91	4
	93	4
	95	2
	112	5
	121	1
	126	3
murrayana	123	5
	131	5
	135	3
bolanderi	136	5

Table 4-2: Sequences of the amplification and sequencing primers used in this study.

	Primer	Sequence 5’ – 3’
Chloroplast	c	CGAAATCGGTAGACGCTACG
	d	GGGGATAGAGGGACTTGAAC
	e	GGTTCAAGTCCCTCTATCCC
	f	ATTTGAACTGGTGACACGAG
	trnF-Rb	CTTGCCAGGAACCAGATTTG
Mitochondrial	NAD1B1F	ATGCCGCCCGTTTCCATTTC
	NAD1C1R	TGCTGCAAAGGGTTAGGGGG
	NAD1B3F	CGGGCGAGTCACTTAAAAGTCAC
	NAD1C3R	TTTTAAGTGACTCGCCCGACC

Table 4-3: Relative positions of single nucleotide polymorphisms and a 5 bp indel in the *trnL* intron in 9 populations of lodgepole pine. The numbers above indicate the position relative to the start of the inton.

	Nucleotide Position					
	92	111	389	400	404-408	420
Consensus	C	T	A	G	TAAAT	G
Individual						
33-3 (*latifolia*)	T	–	–	–	–	–
33-14 (*latifolia*)	–	G	–	–	–	–
36-3 (*latifolia*)	–	–	–	–	INDEL	–
95-3 (*contorta*)	–	–	–	–	INDEL	–
49-11 (*contorta*)	–	–	–	–	INDEL	A
49-18 (*contorta*)	–	–	–	–	INDEL	–
123-3 (*murrayana*)	–	–	–	T	–	–
131-26 (*murrayana*)	–	–	C	–	–	–
135-8 (*murrayana*)	–	–	C	–	–	–

Table 4-4: Relative positions of single nucleotide polymorphisms, a 1 bp indel, and a 26 bp indel in the *trnL/F* spacer in 7 populations of lodgepole pine. The numbers above indicate the position relative to the start of the spacer.

	Nucleotide Position						
	11	91	103	228	240	244	253
Consensus	C	T	G	AATTATTCAATT**G**CAG**T**CCATTTTT			
Individual							
31-5 (*latifolia*)	–	–	–		T	–	
31-15 (*latifolia*)	–	–	–	———— INDEL (26 bp) ————			
141-26 (*latifolia*)	–	–	T		–	–	
143-10 (*latifolia*)	–	–	–		–	G	
145-5 (*latifolia*)	–	–	–		–	G	
154-11 (*latifolia*)	A	–	–		–	–	
135-13 (*murrayana*)	–	INDEL	–		–	–	

Table 4-5: Evolutionary distance within (diagonal) and between the four subspecies of lodgepole pine based on the chloroplast *trnL* intron and the *trnF* spacer using the method of Tajima and Nei (1984).

	latifolia	*murrayana*	*contorta*	*bolanderi*
latifolia	2.09×10^{-4}			
murrayana	1.31×10^{-4}	5.45×10^{-4}		
contorta	3.96×10^{-4}	3.13×10^{-4}	4.93×10^{-5}	
bolanderi	1.06×10^{-4}	2.88×10^{-4}	2.56×10^{-5}	—

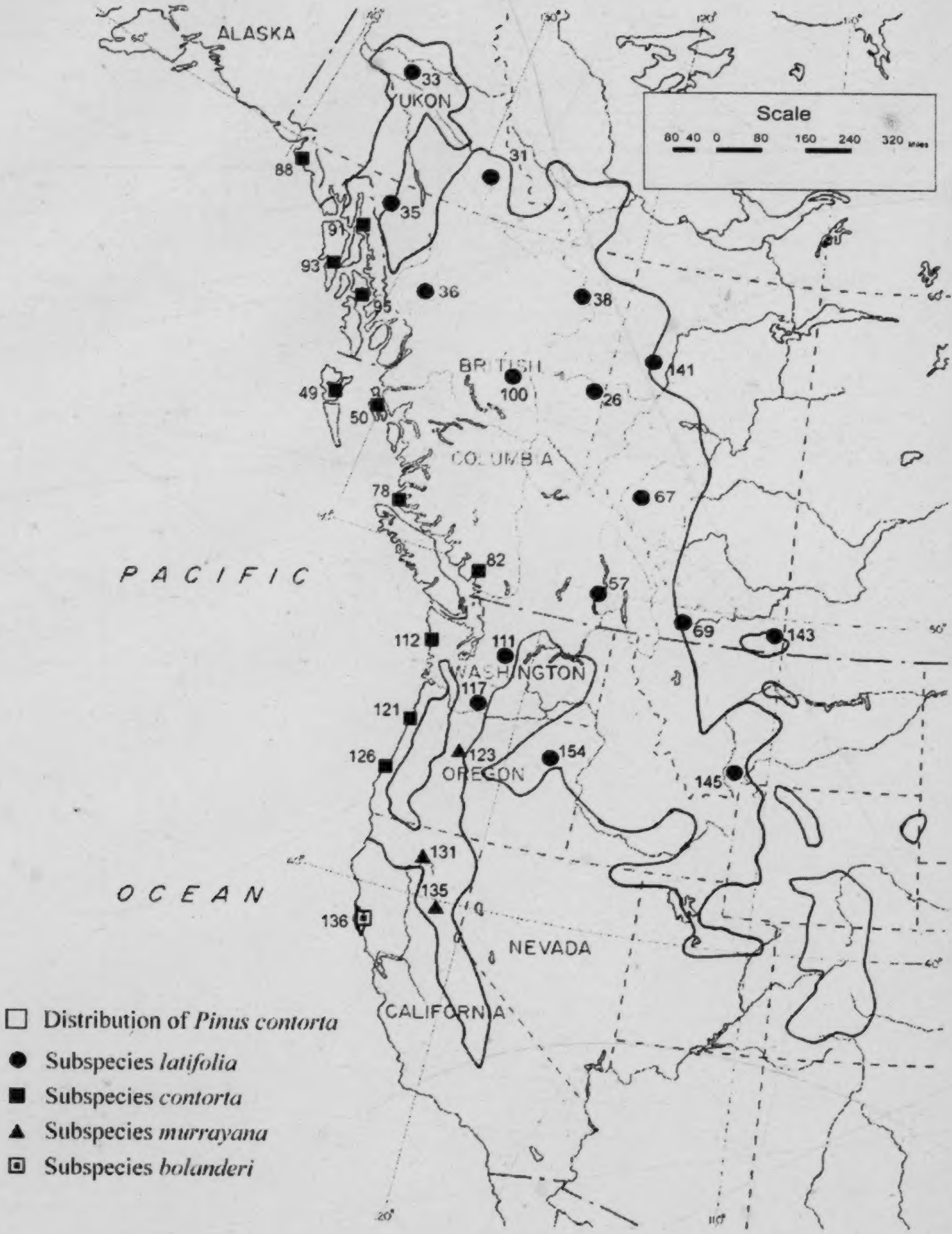
ALASKA
YUKON
BRITISH
COLUMBIA
WASHINGTON
OREGON
NEVADA
CALIFORNIA
PACIFIC
OCEAN
Scale
80 40 0 80 160 240 320 miles
33
31
88
35
91
93
95
36
38
141
100
26
49
50
67
78
82
57
69
143
112
111
117
121
123
154
126
145
131
135
136
□ Distribution of *Pinus contorta*
● Subspecies *latifolia*
■ Subspecies *contorta*
▲ Subspecies *murrayana*
⊡ Subspecies *bolanderi*

Figure 4-1: Locations of 31 populations of *Pinus contorta* used in this study.

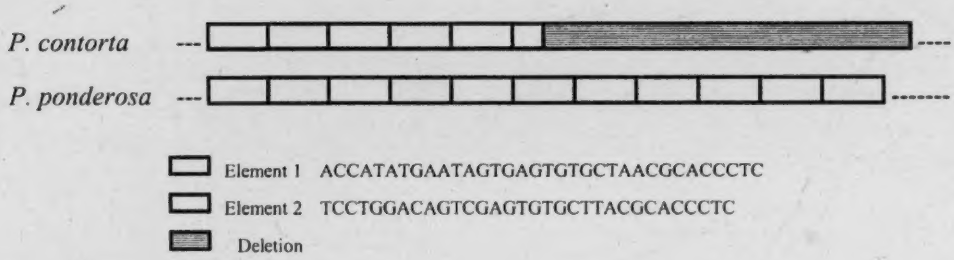
P. contorta
P. ponderosa
Element 1 ACCATATGAATAGTGAGTGTGCTAACGCACCCTC
Element 2 TCCTGGACAGTCGAGTGTGCTTACGCACCCTC
Deletion

Figure 4-2: Comparison of the mitochondrial *nad1* b/c intron repeat region between *Pinus contorta* and *Pinus ponderosa.*

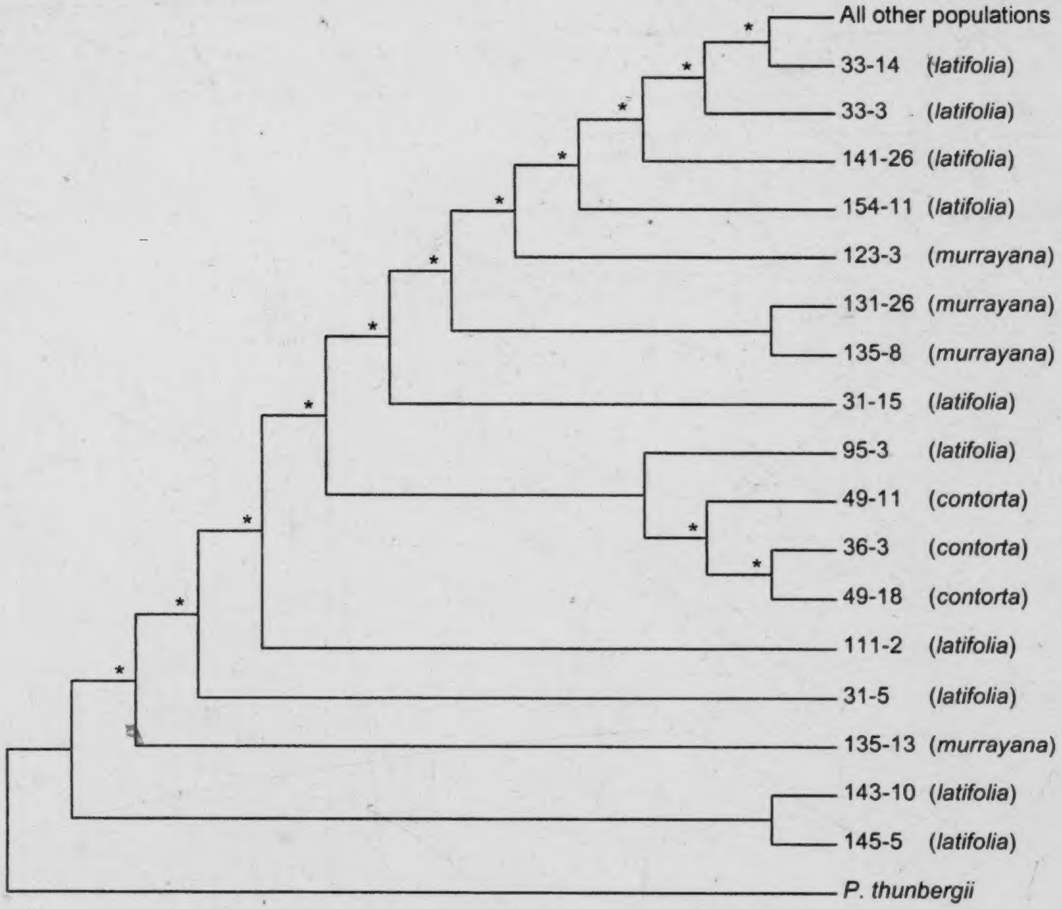
All other populations
33-14 (latifolia)
33-3 (latifolia)
141-26 (latifolia)
154-11 (latifolia)
123-3 (murrayana)
131-26 (murrayana)
135-8 (murrayana)
31-15 (latifolia)
95-3 (latifolia)
49-11 (contorta)
36-3 (contorta)
49-18 (contorta)
111-2 (latifolia)
31-5 (latifolia)
135-13 (murrayana)
143-10 (latifolia)
145-5 (latifolia)
P. thunbergii
*

Figure 4-3: One of 8,610 most parsimonious trees based on the combined chloroplast *trnL* intron and *trnL/F* spacer regions. * indicates branches that collapse in a strict consensus tree.

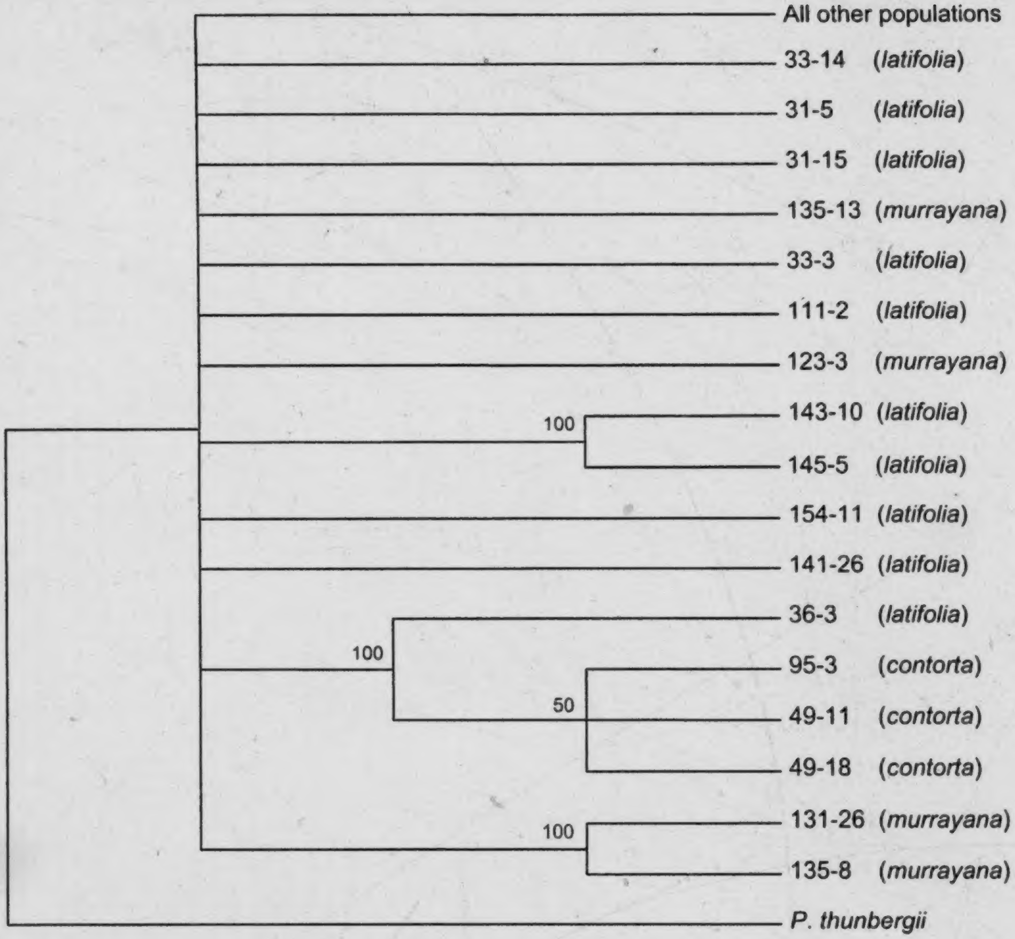
All other populations
33-14 (*latifolia*)
31-5 (*latifolia*)
31-15 (*latifolia*)
135-13 (*murrayana*)
33-3 (*latifolia*)
111-2 (*latifolia*)
123-3 (*murrayana*)
100
143-10 (*latifolia*)
145-5 (*latifolia*)
154-11 (*latifolia*)
141-26 (*latifolia*)
36-3 (*latifolia*)
100
95-3 (*contorta*)
50
49-11 (*contorta*)
49-18 (*contorta*)
100
131-26 (*murrayana*)
135-8 (*murrayana*)
P. thunbergii

Figure 4-4: Majority rule consensus tree of 8,610 most parsimonious trees based on the combined chloroplast *trnL* intron and *trnL/F* spacer regions. Numbers above the nodes indicate the percentage of trees that supported the nodes.

Chapter 5

General Conclusions

Introduction

The various forces affecting populations can be difficult to elucidate and quantify. Although mathematical relationships and models are instrumental in this task, application of models to real data can be challenging. Biological systems are never identical, and although certain patterns do emerge, exceptions can almost always been found for any rule (e.g. Hall 1998). As Lewontin has recently said: "There is no single cause of the evolutionary change in the properties of members of a species. Natural selection may be involved but so are random events, patterns of migration and interbreeding, mutational events, and horizontal transfer of genes across species boundaries. The change in each character of each species is a consequence of a particular mixture of these causal pathways" (Lewontin 2002).

The impact of some of these forces (gene flow and genetic drift, migration, mutation and founder effects) on populations of *Pinus contorta* and the resulting distribution of genetic diversity within and between populations and subspecies has been investigated in this thesis.

Research overview

The research presented in this thesis had three main components. First, an evaluation of the levels of diversity in central, intermediate and marginal populations of *Pinus contorta* subsp. *latifolia* using RAPD markers was conducted (Chapter 2). This addresses one of the long standing theories (Mayr 1970) that marginal populations should have reduced genetic variation compared to centrally located populations. As well as examining estimates of diversity and population differentiation, a multilocus analysis based on the method of Brown *et al.* (1980) and Brown and Feldman (1981) was employed. Despite the availability of these methods, multilocus analyses are often overlooked in studies of genetic diversity. The single locus and multilocus analyses performed here both indicate that marginal populations of subsp. *latifolia* experience reduced variation in comparison to central populations. Whether this is a result of founder effects, or of small population size with limited gene flow from other populations, reduced variability can result in an inability to respond to environmental changes. The multilocus analysis also revealed that associations between loci are dissimilar across populations, indicating that founder effects are an important contributor to the genetic structure of populations of lodgepole pine. This is likely a result of the role of fire in the establishment of populations. Non-random mating, or uneven contribution of seed from parental trees could result in significant initial gametic disequilibria in populations.

The second part of the research (Chapter 3) consisted of an analysis of an expanded number of populations of lodgepole pine, with representatives of each of the four subspecies. The primary objective in this component was to look for evidence of isolation by distance, as originally proposed by Wright (1943) using a model developed by Slatkin (1993). Species with large population sizes and extensive gene flow (such as lodgepole pine) should exhibit such a pattern, with physically closer populations being genetically more similar than physically distant populations. Migration events other than a smooth expansion of the range margin however, can disrupt such a pattern.

The theory of persistence of subsp. *latifolia* in a Yukon refugium during the Wisconsin glaciation has been a topic of much discussion (Hanson 1950; Heusser 1967; von Rudloff and Nyland 1979; Critchfield 1980; Wheeler and Guries 1982; MacDonald and Cwynar 1985, 1991; Cwynar and MacDonald 1987; Yang and Yeh 1995). If populations were able to survive in a northern refugium, genetic divergence from populations existing south of the glacial front can be expected given the reduced sizes of putative northern populations and the long period of separation (~100,000 years). Significant genetic divergence between these groups would disrupt any pattern of isolation by distance that would otherwise be present in current populations.

The analysis of RAPD markers showed that a pattern of isolation by distance is present in *Pinus contorta* subsp. *latifolia*, suggesting that populations in the Yukon have only recently arrived. This conforms well to an analysis of fossil pollen (MacDonald and Cwynar 1985) which suggests that populations of subsp. *latifolia* have reached their northern margin only recently (within the last 100 years) and may still be expanding northward.

Application of the same analysis to populations from subsp. *contorta* and *murrayana* failed to detect any pattern of isolation by distance. For subsp. *contorta*, this also conforms with data from fossil pollen, which suggests that subsp. *contorta* colonized many parts of the coast of British Columbia almost simultaneously (within ~200 years) (Peteet 1991). This suggests that the source of the migrants must have been much closer that the nearest populations south of the glacial front, in the Puget Sound area. Populations on the Queen Charlotte Islands would be the nearest and most likely source of migrants. Although no pattern of isolation by distance was detected in subsp. *murrayana*, this may be a result of the small number of populations sampled.

The third component of the research (Chapter 4) used sequence data to attempt to resolve the four subspecies of *Pinus contorta*. Sequencing of three loci in 140 individuals however, failed to display a pattern of haplotype variation that aligned with the current division of the species. This result may indicate that populations of lodgepole pine have recently expanded from one or a few refugia south of the glacial front, that were invariant at the loci examined. Alternately, the slow rate of change in the genomes studied may be ineffective at determining relationships at the intraspecific level.

Directions for future research

The research completed in this thesis raises additional questions regarding the nature of the relationships between the subspecies of *Pinus contorta*. At least two geographic areas of focus would complement well the body of research into this species to date.

One area that deserves further attention is the Atlin Lake region of British Columbia, and the nearest major river systems to the south. These river systems may be serving as an effective conduit for gene flow between subsp. *contorta* and *latifolia*. The data presented by Wheeler and Guries (1982) and here (Chapter 4) suggest that some degree of introgression of these subspecies is occurring in this area. Study of populations in this region and to the north in the Yukon, may give more evidence to what is suggested here: that the unique nature of populations in the Yukon region is a result of gene flow between these two subspecies, and not due to persistence of populations in a Yukon refugium. Perhaps one of the reasons that populations in this region have not been studied in depth is their remoteness and associated cost of sampling.

Additionally, the zone between subsp. *latifolia* and *murrayana* has not been well researched. This area may afford even more interesting possibilities. Wheeler and Guries (1982) noted that the genetic relationship between the subspecies and the morphological relationship do not align very well. A population in southern Oregon assigned to subspecies *murrayana* based on morphology was genetically more like subsp. *latifolia* than other populations of subsp. *murrayana*. This may be related to the role of cone serotiny (which is a primary determinant of subspecies groups) and selection associated with fire incidence.

Genetic support for the subspecies designations remains elusive. Although they are differentiated morphologically, the RAPD data and the sequencing data of organellar DNA regions presented here did not separate populations into subspecies groups. Sequencing a greater number of organellar loci would be required to yield data that would provide such support. If the paucity of variation detected at the two chloroplast and one mitochondrial loci is an indication of reduced size and number of populations south of the glacial front, then a large number of additional loci may need to be sampled. With greater sampling will come greater certainty regarding the hypothesis of separation of subspecies during the Wisconsin. However, such an effort may not be rewarded. Failure to detect variation does sometimes mean that it does not exist.

References

Brown, A.H.D., M.W. Feldman and E. Nevo. 1980. Multilocus structure of natural populations of *Hordeum spontaneum*. Genetics **96**: 523-536.

Brown, A.H.D. and M.W. Feldman. 1981. Population structure of multilocus associations. Proc. Natl. Acad. Sci. **78**: 5913-5916.

Critchfield, W.B. 1980. Genetics of Lodgepole Pine. U.S. For. Serv. Res. Pap. WO-37.

Cwynar, L.C. and G.M. MacDonald. 1987. Geographical variation of lodgepole pine in relation to population history. Am. Nat. **129**: 463-469.

Hall, B.G. 1998. Adaptive mutagenesis: a process that generates almost exclusively beneficial mutations. Genetica **103**: 109-125.

Hansen, H.P. 1950. Postglacial forests along the Alaska highway in British Columbia. Proc. Am. Philos. Soc. **94**: 411-422.

Heusser, C.J. 1960. Late Pleistocene environments of north Pacific North America. Am. Geogr. Soc. Special Publ. No. 35.

Lewontin, R.C. 2002. Directions in evolutionary biology. Annu. Rev. Genet. **36**: 1-18.

MacDonald, G.M. and L.C. Cwynar. 1985. A fossil pollen based reconstruction of the late Quaternary history of lodgepole pine (*Pinus contorta* ssp. *latifolia*) in the western interior of Cananda. Can. J. For. Res. **15**: 1039-1044.

MacDonald, G.M. and L.C. Cwynar. 1991. Post-glacial population growth rates of *Pinus contorta* ssp. *latifolia* in western Canada. J. Ecol. **79**: 417-429.

Mayr, E. 1970. Populations, Species and Evolution. Harvard University Press, Cambridge, MA.

Peteet, D.M. 1991. Postglacial migration history of lodgepole pine near Yakutat, Alaska. Can. J. Bot. **69**: 786-796.

Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution **47**: 264-279.

von Rudloff, E. and E. Nyland. 1979. Chemosystematic studies in the genus *Pinus*. III. The leaf oil terpene composition of lodgepole pine from the Yukon Territory. Can. J. Bot. **57**: 1367-1370.

Wheeler, N.C., and R.P. Guries. 1982. Biogeography of lodgepole pine. Can J. Bot. **60**: 1805-1814.

Wright, S. 1943. Isolation by distance. Genetics **28**: 114-118.

Yang, R.-C. and F.C. Yeh. 1995. Patterns of gene flow and geographic structure in *Pinus contorta* Dougl. For. Genet. **2**: 65-75.